AALTO UNIVERSITY
SCHOOL OF SCIENCE

Tuomas Nikoskinen

EXPOSURE ESTIMATION OF INTEREST RATE DERIVATIVES WITH AFFINE TERM STRUCTURE MODELS

Master's thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Technology in the Degree Programme in Engineering Physics and Mathematics.

Espoo, September 8, 2014

Supervisor:

Prof. Ahti Salo

Instructor:

M.Sc. (Econ.) Juha Soini

Aalto University
School of Science

AALTO UNIVERSITY SCHOOL OF SCIENCE P.O. Box 1100, FI-00076 AALTO http://www.aalto.fi	ABSTRACT OF THE MASTER'S THESIS

Author:	Tuomas Nikoskinen		
Title:	Exposure Estimation of Interest Rate Derivatives with Affine Term Structure Models		
Degree programme:	Engineering Physics and Mathematics		
Major subject:	Systems and Operations Research	Minor subject:	Computational Science and Engineering
Chair (code):	Mat-2		
Supervisor:	Prof. Ahti Salo		
Instructor:	M.Sc. (Econ.) Juha Soini		

The financial crisis of 2007–2008 with the failures of prestigious large financial institutions has put the management of counterparty credit risk (CCR) in the spotlight. Enormous over-the-counter (OTC) derivatives market and increases in the regulation of economic capital related to OTC derivatives further highlight the need for quantitative modeling of CCR.

Counterparty credit risk is comprised of two components both of which contain uncertainty: the default likelihood of a counterparty to an OTC derivative contract and the magnitude of loss in the case of a default. Estimation of the latter component, known as exposure, is in the scope of this thesis. Estimation of exposure is intrinsically challenging due to exposure being tightly linked to values of derivatives which are innately uncertain and volatile.

Application of the affine term structure models (ATSMs) to exposure modeling of interest rate derivatives is presented in this thesis. To this end, a framework for exposure modeling is introduced and the theory of ATSMs reviewed. Parameter estimation in ATSMs is presented from the viewpoint of state space models (SSM) because the setup of ATSMs corresponds to that of state space models (SSMs) which enables the use of Kalman filter extensions in the parameter estimation. Two case studies are used to illustrate exposure modeling with a particular ATSM.

This thesis is application-driven which means that the interest is in practical computation of exposure related to interest rate derivatives within the CCR context. The approach is decidedly computational and the focus is on facilitating exposure modeling in practice.

Date:	September 8, 2014	Language:	English	Number of pages:	85

Keywords: Exposure, Counterparty credit risk, Affine term structure models, Derivatives, Continuous-discrete Kalman filter, Parameter estimation

Aalto-yliopisto
Perustieteiden
korkeakoulu

AALTO-YLIOPISTO PERUSTIETEIDEN KORKEAKOULU PL 1100, FI-00076 AALTO http://www.aalto.fi	DIPLOMITYÖN TIIVISTELMÄ

Tekijä:	Tuomas Nikoskinen		
Työn nimi:	Korkojohdannaisten vastapuoliriskialtistuksen estimointi affiineilla korkokäyrämalleilla		
Tutkinto-ohjelma:	Teknillinen fysiikka ja matematiikka		
Pääaine:	Systeemi- ja operaatiotutkimus	Sivuaine:	Laskennallinen tiede ja tekniikka
Opetusyksikön koodi:	Mat-2		
Valvoja:	Prof. Ahti Salo		
Ohjaaja:	KTM Juha Soini		

Finanssikriisi 2007–2008 yhdessä arvostettujen finanssi-instituutioiden vararikkojen kanssa on lisännyt entisestään vastapuoliluottoriskin hallinnan tärkeyttä. Valtavat yksityisesti vaihdettujen johdannaisten markkinat ja lisääntyvä sääntely näiden johdannaisten pääomavaateille luovat tarvetta vastapuoliluottoriskin kvantitatiiviselle mallintamiselle.

Vastapuoliluottoriski koostuu kahdesta epävarmuutta sisältävästä komponentista: yksityisen johdannaissopimuksen vastapuolen konkurssitodenäkköisyydestä ja tappion suuruudesta konkurssin tapahtuessa. Jälkimmäisen komponentin, joka tunnetaan altistuksena, estimointi kuuluu tämän työn aihepiiriin. Altistuksen estimointi on haastava tehtävä, koska tappion suuruus liittyy vahvasti johdannaisen hintaan, joka on luonnostaan volatiili ja sisältää epävarmuutta.

Tässä työssä esitellään affiinien korkokäyrämallien soveltaminen vastapuoliluottoriskiin liittyvän altistuksen estimointiin. Altistuksen estimointia varten työstetään viitekehys ja käydään läpi korkokäyrämallien ja parametrien estimoinnin teoriaa. Affiinit korkokäyrämallit voidaan esittää tilayhtälömalleina, jonka ansiosta näiden korkokäyrämallien parametrien estimointi voidaan tehdä Kalman-suodin laajennuksia käyttäen. Työssä käytetään kahta esimerkkitapausta havainnollistamaan altistuksen estimointia valitulla affiinilla mallilla.

Tämä työ on sikäli sovelluslähtöinen, että työn pääpaino on korkojohdannaisiin liittyvän vastapuoliriskialtistuksen laskennassa. Työssä käytetty lähestymistapa on laskennallinen, ja se tähtää altistuksen estimointiin käytännössä.

Päivämäärä:	8.9.2014	Kieli:	Englanti	Sivumäärä:	85
Avainsanat:	Altistus vastapuoliluottoriskille, affiinit korkokäyrämallit, johdannaiset, jatkuva-diskreetti Kalman suodin, parametrien estimointi				

Preface

This work was started in the Market Risk Group in the Risk Management Department at Pohjola Bank plc, Finland. The work draws heavily on the knowledge and expertise gained from my employment therein. I am very grateful for the opportunity to work in the Market Risk Group and would like to thank Juha Soini and Mauri Larikka for all the guidance, insights and discussions during my employment — it was an honor to work with you.

I wish to express my sincere gratitude to my instructor Juha Soini and supervisor Ahti Salo for their support, trust and guidance during this thesis project. I am also very much obliged to Arno Solin for his unwavering companionship and help — be that with LaTeX, the types of dashes or infinite-dimensional formulas — throughout my studies. Furthermore, I am grateful to Helsingin Osakesäästäjät ry for their financial support for this project.

Finally, this thesis would not have materialized without the support and help from my family and Pauliina. Thank you.

Helsinki, 2014

Tuomas Nikoskinen

Contents

Symbols and Abbreviations

Matrices are capitalized and vectors are in bold type.

Operators and miscellaneous notation

dd.mm.yyyy	Date format, e.g. 6.5.2013 denotes May 6, 2013
$1 : n$	$1, 2, \ldots, n$
$a : b : c$	Points from a to c with constant interval b
$\mathbf{A}^\top$	Matrix transpose
$\mathbf{I}$	Identity matrix
$\mathrm{diag}(\mathbf{x})$	Diagonal matrix with vector $\mathbf{x}$ on the diagonal
$\mathrm{E}(\mathbf{b})$	Expectation of $\mathbf{b}$
$(h)^+$	$\max\{h, 0\}$
j	The imaginary unit
$\mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$	Gaussian distribution with mean $\mathbf{m}$ and covariance $\boldsymbol{\Sigma}$
$p(\mathbf{y} \mid \mathbf{x})$	Conditional probability density of $\mathbf{y}$ given $\mathbf{x}$
$\mathbb{R}, \mathbb{C}$	The real and complex numbers
$\mathbb{R}_+$	The positive real numbers

Symbols

$\mathrm{C}(t)$	Collateral at time t
$\xi(t)$	Exposure at time t
$\xi_C(t)$	Collateralized exposure at time t
$\mathrm{EE}(t)$	Expected exposure at time t
$\mathrm{PFE}^\alpha(t)$	Potential future exposure at time t with confidence level α
$P(t, T)$	Time-t price of a zero-coupon bond with maturity at T

Abbreviations

ATSM	Affine term structure model
AFNS model	Affine arbitrage free Nelson-Siegel model
CCR	Counterparty credit risk
CKF	Cubature Kalman filter
IRS	Interest rate swap
MAP	Maximum a posteriori
ODE	Ordinary differential equation
OTC	Over-the-counter, i.e. privately negotiated
SDE	Stochastic differential equation
SSM	State space model
ZCB	Zero-coupon bond

1 Introduction

The basis for all risk management lies in the identification, assessment and quantification of risks. Due to the intrinsic uncertainty in risks, both in terms of the occurrence likelihood and the magnitude of consequences, probabilistic inference is favored in *quantitative risk analysis* (Ayyub, 2003; McNeil *et al.*, 2005). The approach in quantitative risk analysis, roughly, is to estimate separately the risk occurrence likelihood and the magnitude of loss that would result from a realized risk. The latter aspect is generally referred to with *exposure* estimation.

The financial crisis of 2007–2008 with the failures of prestigious large financial institutions such as Lehman Brothers and AIG has put financial risk management and especially the management of *counterparty credit risk* (CCR, Pykhtin, 2005; Gregory, 2010) in the spotlight. Counterparty credit risk can be thought of as credit risk between counterparties to a privately traded (over-the-counter, OTC) derivative contract. An OTC derivative is a financial agreement that derives its value from prevailing values of underlying items, such as interest rates or assets, that have no intrinsic value. In addition to the recent financial calamities, the assessment of CCR is further motivated by the associated regulatory capital requirements, that are based on estimates of possible loss, and enormous size of the OTC derivatives market; the estimated global market size for OTC derivatives, measured in notional principal, exceeded that of exchange traded derivatives by a ratio of more than $10 : 1$ in 2012[1]. The vast majority of all OTC derivatives are *interest rate derivatives*, which represent almost 80 % of the total notional. Hence, we restrict our attention to this derivative type hereafter.

Assessing counterparty credit risk quantitatively is complex; the risk stems from a possible counterparty default, which would cause a loss that is tightly connected to the innately volatile and uncertain values of derivatives. Default likelihood estimation associated with this credit risk is typically tackled with credit-rating based systems and models. These aspects are not considered in this thesis. Rather, we focus on the other part of quantitative

[1] See `www.bis.org/statistics/r_qa1306_hanx23a.pdf` for exchange traded derivatives and `www.bis.org/statistics/dt1920a.pdf` for OTC derivatives market size.

CCR analysis, that is exposure estimation. The reason why exposure estimation is so challenging is the aforementioned link to derivatives' values. While with the traditional credit risk, that is in simplistic terms born from one party lending money to another, the amount of possible loss at future time instances (exposure) is known fairly precisely, this is not the case with counterparty credit risk. The value of a derivative varies in time, because the underlyings vary, and due to leverage, a small change in the underlying has typically a large impact on a derivative value.

At the heart of the exposure estimation problem is the need to estimate future values of derivatives from which exposure estimates can be obtained. Pykhtin and Zhu (2007) present a blueprint for modeling exposure that is founded on a simulation based approach — a similar approach was earlier introduced in more general level for CCR modeling by Canabarro and Duffie (2003). The blueprint consists of scenario generation and pricing steps; future scenarios for relevant risk factors are generated using a simulation model and the derivatives in question are valued in each scenario given the simulated risk factor values. Stochastic differential equations (SDEs, Karatzas and Shreve, 1991; Øksendal, 2003) are therein referred to as a popular method for future scenario generation. Significantly more extensive, sophisticated and detailed framework for exposure modeling is presented by Cesari *et al.* (2009). The underlying simulation based approach is, in principle, still the same. In this thesis exposure modeling is built on the framework of Pykhtin and Zhu (2007).

In the literature on exposure modeling of interest rate derivatives, the body of research focuses on either pricing given the present time market conditions or econometric analysis of the *yield curve* and its time-evolution; a yield curve is a mapping of different maturities to corresponding interest rates, yields. Pricing derivatives based on simulated risk factor values has almost solely concentrated on zero-coupon bond pricing, because these prices are interchangeable with yields. Research on quantitative exposure modeling from a computational viewpoint is rare according to our best knowledge; either the focus is in exposure measure analysis, as in Zhu and Pykhtin (2006), or exposure modeling has been used as a demonstration of pricing model applications, see e.g. Singleton and Umantsev (2002). An exception to this is provided by Cesari *et al.* (2009) who consider rigorous quantitative

exposure modeling at a computation level.

The amount of literature related to interest rate derivatives and their pricing is immense. We highlight here a few references based on relevance with regards to the topic of this thesis. Brigo and Mercurio (2006) provide the *de facto* review of interest rate term structure models and their application to pricing. Shreve (2004) presents the fundamental stochastic mathematics behind derivative pricing in more advanced technical level. Glasserman (2003) addresses the intersection of Monte Carlo methods and derivatives pricing. Singleton (2006) focuses on the interplay between financial theory and empirical study of dynamic asset pricing.

It is worth noting that after the 2007–2008 financial crisis the market practice for derivative pricing has evolved towards the use of multi-curve pricing (Pallavicini and Tarenghi, 2010), in which relevant forward rates and discount factors are computed from different curves. The classical single curve based pricing can be seen as a simplification of the multi-curve based pricing causing notable problems only in instances where forward rates from different forward curves are needed. There are complicated issues regarding multi-curve pricing theory which we do not address in this thesis.

Affine term structure models (ATSMs, Duffie and Kan, 1996; Dai and Singleton, 2000) is the class of models applied in this thesis to exposure modeling. This choice is motivated by the flexible specification of ATSMs and the availability of closed form solutions for relevant fundamental quantities within this model class. ATSMs build on the concept of riskless short rate and extend it to multiple dimensions — one dimensional short rate based term structure modeling was introduced by Vasicek (1977) and Cox *et al.* (1985). Additionally, it is possible to specify an ATSM with a jump process that would introduce discontinuous jumps to the short rate process (see, e.g., Duffie *et al.*, 2000). This feature is not considered in this thesis. Moreover, theoretical advances with ATSMs do not yet provide for multi-curve pricing and hence they are considered in the classical single curve pricing framework.

Practical use of an ATSM requires estimation of the model parameters, or *model calibration* as it is often called. Because the setup of ATSMs correspond to that of state space models (SSMs, Durbin and Koopman, 2012),

the celebrated Kalman filter (Kalman, 1960; Jazwinski, 1970), and its descendants, known from state estimation problem within the SSM context, can be utilized in parameter estimation of ATSMs. This greatly facilitates computations related to model calibration. Jong (2000) demonstrates application of the Kalman filter in parameter estimation of ATSMs. Lund (1997) approaches the same problem using zero-coupon bond prices as observations instead of yields, which basically makes the problem nonlinear. Comparison of different nonlinear Kalman filters in the state estimation problem, that is tightly connected to parameter estimation, is presented in Christoffersen *et al.* (2012). For parameter estimation in more general SDE environment using Kalman filter extensions, see Mbalawata *et al.* (2013). Kalman filter based parameter estimation is the chosen approach in this thesis.

This thesis is application-driven, which means that the interest is to show how to compute exposure related to interest rate derivatives within the counterparty credit risk context in practice. In the first chapter, we consider exposure modeling generally. The next step is to introduce and review affine term structure models, which are applied to exposure modeling. Linear and nonlinear parameter estimation in state space models using continuous-discrete Kalman and cubature Kalman filters, respectively, is presented next. These methods are directly applicable for calibration of ATSMs. Finally, exposure modeling is illustrated with two case studies using the affine arbitrage free Nelson–Siegel (AFNS) model of Christensen *et al.* (2011), which is an affine term structure model with specific parametrization. Relevant formulas are derived on implementation level and the model is applied to exposure modeling of a swap and a portfolio that consists of an interest rate cap and swap. These case studies are discussed in some detail.

The main contribution of this thesis is to show how exposure of interest rate derivatives can be estimated using affine term structure models in the classical single curve pricing framework. The approach is decidedly computational and the focus is on facilitating exposure modeling in practice. Regarding existing literature, the computational side here is far more detailed than in Pykhtin and Zhu (2007) and focused on interest rate derivatives, in contrast to the treatment of Cesari *et al.* (2009) who accommodate exposure modeling across different derivative types.

2 Exposure in Counterparty Credit Risk Context

We define *counterparty credit risk* (CCR, Pykhtin, 2005; Gregory, 2010) in the words of Gregory:

> "*Counterparty credit risk is the risk that a counterparty to a derivatives transaction will default before expiration of the trade and will not therefore make the current and future payments required by the contract.*"

Derivatives are financial contracts that *derive* their value from one or more underlyings, which can be market variables or something else. Derivatives subject to counterparty credit risk are privately negotiated — commonly known as over-the-counter (OTC) derivatives — because the current and future cash flows determined by such contracts are lost in the event of a contract counterparty default. This is contrary to the case with non-privately traded, that is exchange-traded, derivatives, where the contract cash flows are guaranteed by the exchange and therefore no counterparty credit risk is involved.

Characteristic features of counterparty credit risk are time-dependent, highly uncertain *counterparty credit exposure* and bilateral nature of the risk; credit exposure refers to the magnitude of loss that would result from a counterparty default and bilateral risk to both counterparties having risk to the other. Uncertainty in the exposure stems from fluctuation in the present value of the future cash flows determined by a derivative contract. Moreover, the net value can be either positive or negative resulting in a bilateral risk, because either counterparty can default incurring, possibly, a loss on the other counterparty — whether a loss is incurred or not depends on the default time value of the remaining cash flows. These characteristics of counterparty credit risk set it apart from traditional credit risk, where the exposure is usually known with a high degree of certainty at any time and the risk is unilateral.

Evaluating counterparty credit risk in a quantitative manner requires one to assess both the likelihood of counterparty default and the magnitude of counterparty credit exposure — this setup corresponds with a classical risk

analysis approach (see, e.g. Modarres, 2006) applied in the CCR context. We assume that the default likelihood and exposure are independent of each other. This assumption makes it possible to study them separately enabling us to solely focus on exposure estimation. Note that in circumstances where the independency assumption is a non-valid simplification, one also needs to consider *wrong-way risk*, which refers to a dependency between the default likelihood and exposure that increases counterparty credit risk, i.e. exposure is higher when default is more likely (De Prisco and Rosen, 2005).

2.1 Counterparty Credit Exposure

In the context of counterparty credit risk, the loss resulting from a counterparty default on derivative contracts is known as counterparty credit exposure (see, e.g., Cesari *et al.*, 2009), hereafter referred to as exposure:

Exposure is the magnitude of loss conditional on counterparty default.

The amount of loss is determined by the default time value of future cash flows of the derivative contracts in question — these cash flows have not been received and will not be due to the default. Because the time-t value (price) of a derivative equals that of its future cash flows, exposure is directly coupled with derivatives' values. For example, in the case of a single derivative contract, exposure equals the present value of the derivative contract if it is positive; otherwise it is zero — negative exposure would imply a gain through default which is not possible.

The relation between exposure and values of derivatives implies that exposure shares the time-dependent and stochastic nature of derivatives. The time of a possible default event is not known beforehand and therefore, exposure is relevant throughout a derivative's period of validity. Furthermore, exposure is dependent on the default time values of derivatives, which determine the amount of possible loss. The value of a derivative is determined by its underlyings at any given time, including the default time. It follows from this that exposure is both time-dependent and stochastic because the possible default time is not known and the underlyings, for example interest rates or commodities, are uncertain in the future.

To formulate exposure more rigorously, we denote the value of an OTC derivative contract i, at a future time instance t, with $V_i(\mathbf{u}, t) : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}$, where $\mathbf{u}(t) \in \mathbb{R}^n$ is a random variable representing the underlyings of the derivative. Hence, the contract level exposure $\xi_i \in \mathbb{R}_+$, at time t, is a scalar random variable given by

$$\xi_i(t) = \max\{V_i(\mathbf{u}, t), 0\}. \tag{1}$$

At present time t_0 the values $\mathbf{u}(t)$ are known and the corresponding exposure, often referred to as *current exposure*, is deterministic. Due to the stochastic nature of exposure, our principal interest lies in the time-varying probability distribution of exposure $p(\xi_i, t)$.

The reasoning behind the exposure formula (1) is convenient to explain in terms of replacement cost, as is done in Pykhtin and Zhu (2007). Let us assume that we have previously struck an OTC derivative contract with a counterparty that now defaults. If the contract value is negative for us, we pay it to the defaulting counterparty. But we could receive an equal positive value from the market — assuming that active markets exists for the kind of derivative — if we decided to enter into a similar contract at the time. Therefore, our net loss is defined to be zero in this case. On the other hand, if the contract value is positive when the default happens, we receive nothing from the defaulting counterparty while the price of acquiring a similar contract from the market is the current value of the contract. In this case, our net loss is equal to the value of the contract. Note that the loss can be determined this way even if the non-defaulting counterparty would not replace the contract. Furthermore, by convention, a zero *recovery-rate* is assumed in all exposure computations (Basel Committee on Banking Supervision, 2006), which means that zero percentage of the loss is assumed to be recovered.

To generalize the contract level exposure to a counterparty level exposure, possible *netting agreements* need to be accounted for. A netting agreement defines a set of derivatives, *netting set*, the values of which are aggregated in the case of a default. Exposure of the derivatives belonging to a netting set is simply the maximum of zero and the net value of these derivatives. Without a netting agreement the exposure of multiple derivatives is a sum of

contract level exposures. Counterparty level exposure $\xi(t) \in \mathbb{R}_+$ is therefore

$$\xi(t) = \max\left\{\sum_{i \in \text{NS}} V_i(\mathbf{u}, t), 0\right\} + \sum_{i \notin \text{NS}} \max\{V_i(\mathbf{u}, t), 0\}, \tag{2}$$

where NS denotes the netting set. The first and second term in formula (2) consist of the derivatives that belong and do not belong to the netting set, respectively. Multiple netting sets can be treated with similar logic.

2.1.1 Time-Dependent Exposure Measures

In the following, we define two popular time-dependent risk measures used to quantify exposure in counterparty credit risk context. We follow definitions of the Basel Committee on Banking Supervision (2006). Time-dependent exposure measures quantify exposure in a dynamic manner throughout the period from the present time till a contract's maturity (or the longest maturity in a portfolio). Exposure measures that are not time-dependent (time-invariant exposure measures) can be derived from the time-dependent ones.

Expected exposure (EE) and *potential future exposure* (PFE) are the two core measures used to quantify exposure. Both are time-dependent. EE is the time-varying expectation of exposure (2)

$$\text{EE}(t) = \text{E}^{\mathcal{P}}[\xi(t)], \tag{3}$$

which is taken with respect to the physical probability measure[1] $\mathcal{P}$, that is also known as the actual, real-world or historical probability measure. Under this measure probability distributions are based on historical observations of relevant quantities — contrary to the case with the $\mathcal{P}$ measure, probabilities under the risk-neutral pricing measure $\mathcal{Q}$ are defined through the no-arbitrage theory and present market conditions (Harrison and Kreps, 1979; Harrison and Pliska, 1981). PFE is an estimate of how high the exposure could potentially be at a given time with chosen confidence level.

[1]There is some controversy about the probability measure under which the exposure measures should be defined, see, for example, the discussion on pp. 91–93 in Gregory (2010). From risk management perspective, the choice should be the physical measure, as defined here. However, on p. 261 in Basel Committee on Banking Supervision (2006), it is stated that the supervisors do not restrict a practitioner to employ the physical measure due to the issue of feasibility. That is, the risk-neutral measure can also be used.

Technically, PFE is a quantile and, in that sense, similar to value-at-risk (VaR, McNeil *et al.*, 2005). PFE is defined by

$$\mathrm{PFE}^{\alpha}(t) = \inf\{y \in \mathbb{R} \mid p(\xi(t) \leq y) \geq \alpha\}, \tag{4}$$

where α is a chosen confidence level and the probability $p(\xi(t) \leq y)$ is taken under $\mathcal{P}$. Note that the expected exposure of a portfolio, with several counterparties, is the sum of all counterparty level EEs where as for portfolio level PFE, the corresponding exposure distribution needs to be computed first and only after that apply (4).

The aforementioned exposure measures lend themselves conveniently to visual illustration. Figure 1 shows the exposure concept and measures with the help of an arbitrary derivative contract. Panel (a) shows a set of time-varying scenarios that demonstrate a possible evolution of a derivative's market value in time — four sample scenarios are denoted in black and a number in light grey. Contract market values at time t' are highlighted in red and an approximative continuous distribution of these is shown in panel (b), where $\mathrm{EE}(t')$ and $\mathrm{PFE}^{95\%}(t')$ are denoted with vertical dashed lines. In panel (c) is shown how $\mathrm{EE}(t)$ and $\mathrm{PFE}^{95\%}(t)$ of this contract would evolve in time given the market value scenarios and the corresponding exposures depicted in gray — a plot of $\mathrm{EE}(t)$ or $\mathrm{PFE}^{\alpha}(t)$ is commonly known as *exposure profile* of a contract or portfolio.

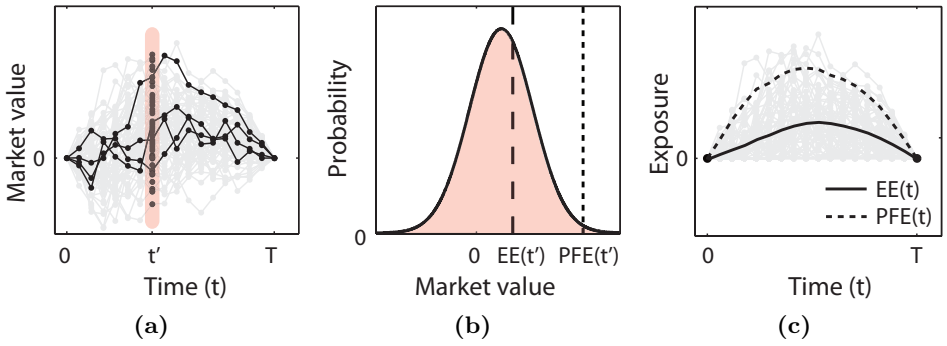


Figure 1: Panel (a) shows a set of simulated scenarios of an arbitrary derivative contract's market value evolution. An approximative continuous probability distribution of the market values at time t' is shown in panel (b). Panel (c) shows how $\mathrm{EE}(t)$ and $\mathrm{PFE}(t)$ evolve in time given the market value scenarios' exposures that are depicted in light gray on the background.

2.1.2 Time-Invariant Exposure Measures

Expected positive exposure (EPE) and *effective expected positive exposure* (effective EPE) are the two time-invariant exposure measures that are involved in the counterparty credit risk related regulatory capital requirements (see, Basel Committee on Banking Supervision, 2006). For this reason, we present these two static measures and their relation to the time-dependent expected exposure. The review is based on the definitions given in Basel Committee on Banking Supervision (2006), and a similar one can be found for example in Gregory (2010).

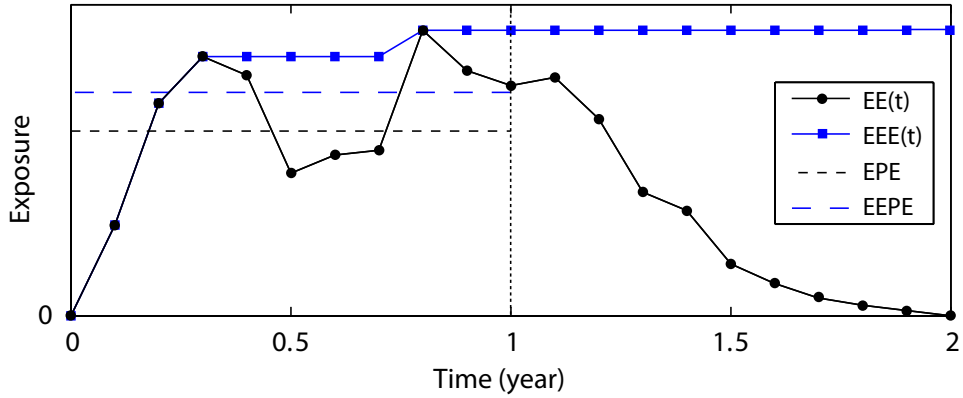


Figure 2: Expected exposure (EE) and the derived exposure measures. Effective EE (EEE) is expected exposure that is constrained to be non-decreasing. Expected positive exposure (EPE) and effective EPE (EEPE) are obtained from EE and EEE by taking average of them over the first year, respectively.

Figure 2 shows an arbitrary expected exposure (EE) profile and three exposure measures derived from it. The *effective expected exposure* (EEE) is obtained from the expected exposure by constraining it to be non-decreasing. EPE and effective EPE (EEPE), the static measures of interest, are the first year averages of EE and effective EE, respectively.

EPE and Effective EPE can be computed by taking the following averages

$$\text{EPE} = \sum_{t_k \leq 1\text{yr}} \text{EE}_k \, \Delta t_k$$
$$\text{Effective EPE} = \sum_{t_k \leq 1\text{yr}} \text{EEE}_k \, \Delta t_k,$$

where Δt_k is the year fraction $t_k - t_{k-1}$ between the two following time

instances on which exposure is computed. The expected exposure at time t_k is denoted with $\mathrm{EE}_k \triangleq \mathrm{EE}(t_k)$ and EEE_k is the effective expected exposure, at time t_k. The effective expected exposure can be computed recursively by

$$\mathrm{EEE}_k = \max\{\mathrm{EE}_k, \mathrm{EEE}_{k-1}\}.$$

2.2 The Impact of Collateral

A *margin agreement* (see, e.g., Pykhtin, 2009; Gregory, 2010) is a contract used to mitigate counterparty credit risk in OTC derivative trades. In principle, the agreement empowers the counterparty having a positive exposure to call for *collateral* from the other counterparty — in case of a default, the collateral will offset part of the resulting losses. In *two-way agreements* both counterparties are eligible to call for collateral, where as in *one-way agreements* only the other counterparty receives collateral. Because exposure varies in time, the amount of collateral needs to be re-adjusted periodically to reflect the changes in exposure; additional collateral is required if exposure increases and *vice versa* (see, Cesari *et al.*, 2009, for a practical example). Following components define the mechanics of margin agreements:

Threshold: The level of exposure below which no collateral is required. For exposures higher than the threshold the difference between these two is required as a collateral.

Initial Margin: The amount of collateral posted upfront. It is independent of any collateral adjustments during the contract lifetime.

Margin Call Frequency: Time-interval between calls to re-adjust the amount of collateral.

Minimum Transfer Amount: Minimum amount of collateral that is called for; calls for smaller collateral adjustment are ignored.

In principle, accounting for collateral in exposure computation is straightforward. The collateralized exposure $\xi_C(t)$ at time t is the difference between uncollateralized exposure $\xi(t)$ and the amount of collateral $C(t) \in \mathbb{R}$

$$\xi_C(t) = \max\{\xi(t) - \mathrm{C}(t), 0\}. \tag{5}$$

The challenge in computing the collateralized exposure is to determine the amount of collateral at margin call dates. This can be either computationally very demanding, as in the case of a margin agreement featuring a minimum transfer agreement and a high margin call frequency, or require additional modeling effort, because securities or foreign currency, instead of cash in base currency, is used as collateral and as a consequence, the collateral value is stochastic. Furthermore, there is a time lag between the latest delivery of collateral and the time when a loss is realized in the case of a default. Pykhtin (2009) calls this time lag a *margin period of risk* and incorporates it explicitly into the calculation of collateralized exposure.

Let us consider the amount of collateral $C_A(t_k)$, at a margin call date t_k, in a two-way margin agreement between counterparties A and B. The margin agreement covers a derivatives portfolio transacted between the two counterparties; we denote the market value of this portfolio, from the point of view of counterparty A, with $V_P(t_k) = \sum_i V_i(t_k)$, where $V_i(t_k)$ is the value of a single derivative at time t_k. As in Cesari *et al.* (2009), the amount of collateral exchanged $\Delta \mathrm{C}_A(t_k)$ — positive if A receives it and *vice versa* — at a marginal call date t_k, is

$$\Delta \mathrm{C}_A(t_k) = \max\{V_P(t_k) - H_B,\, 0\} - \max\{H_A - V_P(t_k),\, 0\} - \mathrm{C}_A(t_{k-1}), \quad (6)$$

where t_{k-1} is the previous margin call date and thresholds for the counterparties are $H_B \geq 0$ and $H_A \leq 0$. The total amount of collateral is then

$$\mathrm{C}_A(t_k) = \mathrm{C}_A(t_{k-1}) + \Delta \mathrm{C}_A(t_k) \cdot 1_{|\Delta \mathrm{C}_A(t_k)| \geq m}, \quad (7)$$

where m denotes the minimum transfer amount and $1_{|\Delta \mathrm{C}(t)| \geq m}$ is an indicator function taking value one if the argument condition holds and zero otherwise. If securities, or cash in foreign currency, are used as collateral instead of cash in base currency, the previous collateral amount $\mathrm{C}_A(t_{k-1})$ needs to be revalued at time t_k.

If the margin agreement is specified as a one-way agreement so that only counterparty A is eligible to call for collateral, the formula (6) for collateral exchange takes the form

$$\Delta \mathrm{C}_A(t_k) = \max\{V(t_k) - H_B,\, 0\} - \mathrm{C}_A(t_{k-1}),$$

and the formula (7) for the total amount of collateral remains unchanged. This way specified one way collateral exchange implies that collateral is received back during the contract if exposure decreases.

Finally, consider the case of computing collateralized exposure in the presence of a daily margin call frequency (that is, according to Gregory, 2010, a market standard) and a non-zero minimum transfer amount. This margin agreement setup implies that the exposure should be evaluated daily in order to compute the amount of collateral to be exchanged — this is computationally extremely intensive task and generally not feasible. To cope with a high margin call frequency, Pykhtin (2009) shows how an approximative collateralized exposure can be computed.

2.3 Framework for Modeling Exposure

We conclude this chapter by presenting a general framework for the estimation of counterparty credit exposure. The framework provides a conceptual summary of the computational steps involved in the exposure estimation. It is defined in more specific terms for exposure estimation of interest rate derivatives that are traded in a single base currency. A similar blueprint for calculating exposure is presented by Pykhtin and Zhu (2007).

The objective in exposure modeling is to obtain the time-dependent measures, expected exposure $\mathrm{EE}(t)$ and potential future exposure $\mathrm{PFE}(t)$, that characterize exposure over time. These measures are easily calculated for a chosen set of dates, referred to as *exposure dates* hereafter, once the corresponding exposure distributions are known — one could say that obtaining the time-varying exposure distribution is the ultimate goal in exposure modeling, however, in practical applications, point measures are typically needed for decision making. Exposure dates include the present time and a set of, often unequally spaced, dates till a contract's (or a portfolio's longest contract's) maturity. Typically, spacing of the exposure dates is denser near the present time, say an interval of a week or a month, and more sparse further in the future, where it could be from a quarter of a year to even five years in the case of very long contracts, for example in a 30 year swap. Sparser exposure date spacing is mainly motivated by computational considerations.

Computing the exposure distribution for chosen exposure dates requires modeling of derivatives' value distributions in the future, wherein lies the difficulty in exposure modeling: how to obtain an estimate of the distribution (under $\mathcal{P}$ measure) of a derivatives's price (computed under $\mathcal{Q}$ measure) on exposure dates in the future. We refer to value distributions, because the value of a derivative is uncertain and distributions characterize this uncertainty.

A simulation based approach to modeling derivatives' future value distributions consists of two main computational steps, simulation and pricing:

1 **Risk factor simulation:** Simulate multiple risk factor scenarios for the chosen exposure dates under the $\mathcal{P}$ measure. That is, use a model calibrated to historical observations to simulate the time evolution of the factors on which the prices of derivatives depend.

2 **Derivative pricing:** Value the derivatives in each risk factor scenario under the $\mathcal{Q}$ measure. That is, determine the price of derivatives in each simulated scenario for each exposure date according to the no-arbitrage principles assuming that the scenario's risk factor values will be that date prevailing real market conditions.

The outcome of these two steps is a set of derivatives' market values on each exposure date. These values are regarded as samples from the corresponding value distributions and can therefore be used to approximate the distributions. The exposure measures are typically computed straight from the samples by converting them to exposures first and then applying the relevant formulas.

Any simulation or pricing requires prior calibration of the models; by model calibration we refer to model parameter estimation. The calibration step is critical, because it has such a strong effect on model behavior. The reason why it is not regarded as a principal step in exposure estimation is that it does not need to be repeated on each new date or each time a exposure estimate for a new trade needs to be obtained — unlike in the case of front office pricing, the main interest in exposure estimation lies in the future value distributions, not in the exact present time fair prices. The calibration is conducted with the help of present time and historical observations of market variables.

It is of pivotal importance that the simulation model produces flexible enough scenarios for the risk factors, while its dynamics is fairly consistent with the observed risk factor history. The variety of simulated scenarios impacts the exposure distributions, especially the tail risk and the likelihood of more extreme events. Even so the scenarios, and the difference between consecutive scenarios, should be plausible enough in regard of the observed history and common sense. Yet, too constrained scenario generation leads to too conservative exposure estimates; particularly the potential future exposure estimate is not really credible with too homogenous scenarios. The pricing model should feature closed form pricing formulas to facilitate efficient pricing.

Regardless of the model or amount of scenarios, the task of estimating exposure could be significant, if the types of relevant derivatives' underlyings span several different asset classes like interest rates, foreign exchange, commodities or credit. The dimensionality of the problem and the computation burden grow with the number of risk factors, not to talk about the challenge of generating consistent joint scenarios for multiple different kinds of risk factors. Harsh simplifications may be needed to tackle extensive exposure estimation in a feasible manner.

In this thesis we focus on exposure estimation of a single currency interest rate derivatives within the classical single curve pricing framework, in which the discount factors and forward rates are computed from the same yield curve (for multi-curve pricing see, e.g., Bianchetti, 2010; Pallavicini and Tarenghi, 2010). The single currency assumption relieves us from considering foreign exchange rates. The single curve pricing is used because the theoretical advances with affine term structure models (ATSMs), which are used for simulation and pricing, do not allow yet for multi-curve pricing.

All interest rate derivatives derive their price from quantities related to interest rates and associated volatilities, which are the risk factors for these derivatives. Instead of modeling these risk factors and their time-evolution directly, we model evolution of the *state variables* that are assumed to characterize the risk factors and their variation with fewer dimensions. The state variables are transformed to appropriate yields and derivative prices on exposure dates for exposure measure computations.

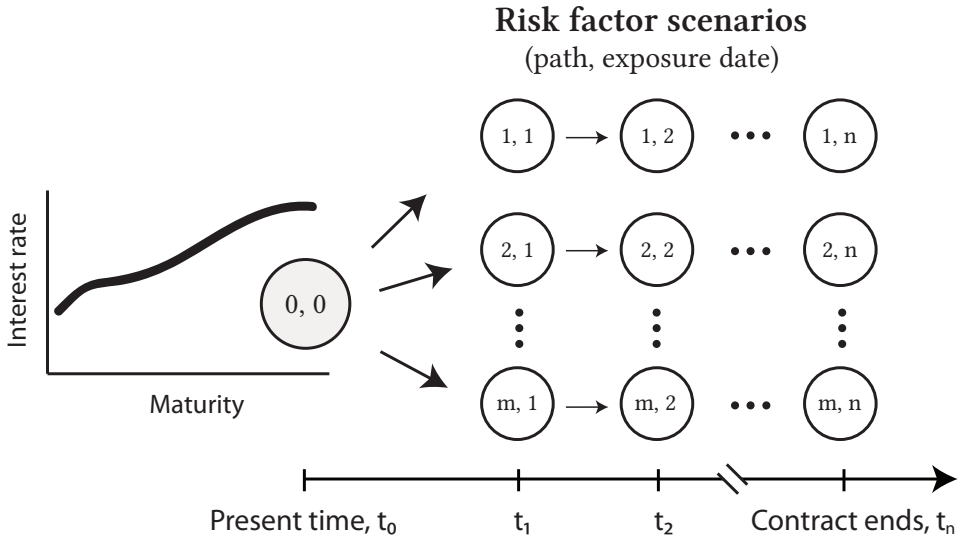


Figure 3: A conceptual visualization of a risk factor simulation for an interest rate derivative. Nodes denote state variables and the present time yield curve is spanned by the estimated present time state variables (shaded node). Future scenarios for the yield curve are obtained through the simulated future state variables, that are illustrated with the nodes on the right.

A conceptual visualization of a risk factor (state variable) simulation for an interest rate derivative is shown in Figure 3. Before the simulation, the model is calibrated to historical market variables through which the present time, and also historical, estimates for the state variables are obtained. Risk factor scenarios are generated in a path-wise manner for the chosen exposure dates. That is, on each path the previous exposure date scenario is the starting point for the next simulation step, by simulating the state variables onwards from the present time estimated state variables.

After the simulation step the derivative in question is priced in each simulated scenario. The outcome of this is a set of possible market values on each exposure date, or from another point of view, paths of market value evolution, of which there is an example in Figure 1a. The market values are transformed to exposures with formula (1) and the exposure measures are computed from these samples. For example, the expected exposure (3) simply is the mean of the exposure scenarios on each exposure date.

3 Affine Term Structure Models

Affine term structure models (ATSMs, Duffie and Kan, 1996; Dai and Singleton, 2000) can be viewed as multi-dimensional generalizations of Vasicek (Vasicek, 1977) and CIR (Cox *et al.*, 1985) models which provide the foundation for interest rate term structure modeling. The *term structure of interest rates*, also known as a yield curve or a zero-coupon curve, is the mapping of maturities T into zero-coupon yields, spot interest rates, $Y(t, T)$ at time t. In modelling this mapping, and how it evolves in time, both Vasicek and CIR models build on the theoretical concept of instantaneous riskless short rate[1] $r(t) \in \mathbb{R}$. The short rate is specified as a one-dimensional affine diffusion process that is defined by a stochastic differential equation (SDE). The entire yield curve can be deduced from the short rate process with the no-arbitrage theory — through closed-form solutions in the case of Vasicek and CIR models. ATSMs generalize the idea of Vasicek and CIR models by defining the short rate as a linear combination of n state variables $\mathbf{x}_i \in \mathbb{R}$, $i = 1, 2, \ldots, n$, which form the state $\mathbf{x} \in \mathbb{R}^n$ and are each governed by a SDE used in Vasicek or CIR models.

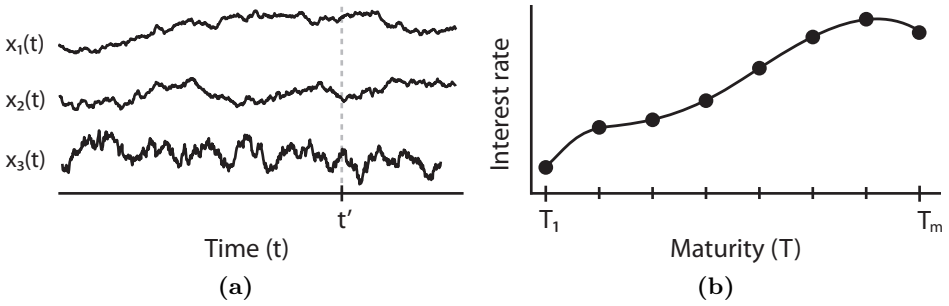


Figure 4: In panel (a) is depicted an example of a term structure model's state variables evolving in time. Panel (b) shows a yield curve solved from the state variables at time t' with maturities from T_1 to T_m.

A consequence of short rate based financial modeling, be that yield curve modeling or option pricing, is that the state driving evolution of the short rate drives the evolution of all model outputs. In essence this means that ATSMs rely on a *latent* and *low-dimensional* state to capture what drives the

[1] Short rate is interpreted as a continuously compounded interest rate at which an entity can borrow money over infinitesimally short period of time.

yield curve and relevant volatilities in the case of derivative pricing; the state is latent (unobservable), because it can only be observed indirectly through measurements, which are yields or prices of derivatives. The difference in the number of dimensions between the state and model outputs, as well as the division of affine term structure models into latent state dynamics and observable outputs, is illustrated in Figure 4. Panel (a) shows examples of state variables' time-evolution; the number of state variables in ATSMs is usually equal to or less than three, i.e. $\mathbf{x} \in \mathbb{R}^n$, $n \leq 3$. In panel (b) is shown a yield curve solved from the time-t' state; typically yields corresponding to roughly dozen maturities, i.e. $m \approx 10$, are modeled and the yields in between these chosen maturities are interpolated.

The aforementioned partitioning of ATSMs into latent state dynamics and observable outputs corresponds exactly to the setup of state space models (SSMs). This similarity gives rise to a state space model interpretation of ATSMs, which makes it easier to estimate parameters and model interest rates and derivatives' prices jointly, because SSMs provide a general probabilistic framework for modeling dynamic systems. A schematic illustration of ATSMs from a state space model viewpoint is presented in Figure 5.

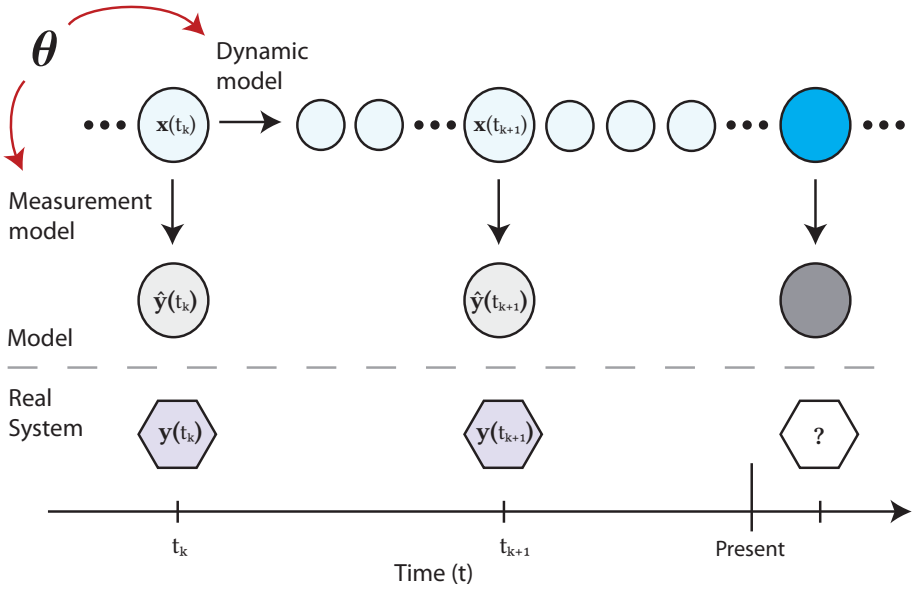


Figure 5: A schematic diagram of ATSMs from a state space model viewpoint. A dynamic model dictates continuous-time evolution of the latent state $\mathbf{x}(t)$ that is transformed into observations $\hat{\boldsymbol{y}}(t)$ at discrete times by a measurement model; parameters $\boldsymbol{\theta}$ affect both of the dynamic and measurement models.

The main aspect of ATSMs not present in the conceptualization of Figure 5 is the need to specify the state dynamics under two *equivalent* probability measures. The equivalence of probability measures means that the measures agree on what outcomes are possible and what are not — but they may assign different probabilities for the possible outcomes. The model dynamics are needed under the risk-neutral probability measure $\mathcal{Q}$, when the model is used for pricing, and under the physical measure $\mathcal{P}$, when the real-world evolution of the model outputs is of interest or historical data is used in parameter estimation. Being able to specify ATSMs simultaneously under two different measures greatly facilitates their use in exposure modeling, because the same model can be used both for risk factor simulation and derivative pricing.

In subsequent sections of this chapter we review the components of which affine term structure models consist. We cover the state dynamics both under the risk-neutral and physical measures and the relation between these two via change of measure. A related topic is the admissibility of the state dynamics which we touch upon briefly. Risk-neutral dynamics give rise to zero-coupon bond and derivative pricing, that are presented in some detail. The chapter concludes with a brief summary of ATSMs from a state space model point of view.

3.1 Risk-Neutral State Dynamics and Zero-Coupon Yields

In affine term structure models the short rate $r(\mathbf{x}, t)$ is specified as an affine function of the n-dimensional state $\mathbf{x}(t) \in \mathbb{R}^n$,

$$r(\mathbf{x}, t) = \rho_0 + \boldsymbol{\rho}_1^\top \mathbf{x}(t), \tag{8}$$

where $\rho_0 \in \mathbb{R}$, $\boldsymbol{\rho}_1 \in \mathbb{R}^n$. For pricing purposes, the state dynamics are defined under the risk-neutral probability measure $\mathcal{Q}$, hence we fix a probability space $(\Omega, \mathcal{F}, \mathcal{Q})$, where the probability measure $\mathcal{Q}$ assigns probabilities to events in the event space $\mathcal{F}$. An event is a set of outcomes belonging to the sample space Ω. Furthermore, we fix an information filtration $\mathcal{F}_t = \{\mathcal{F}_t : t \geq 0\}$ that satisfies the usual conditions (see, e.g., Williams, 1991). The information filtration $\mathcal{F}_t$ can be thought to contain the information about

the state available at time t.

The dynamic model describing time-evolution of the state $\mathbf{x}(t)$ is written in the form of an Itô stochastic differential equation[2] (SDE)

$$\mathrm{d}\mathbf{x}(t) = \mathbf{f}^{\mathcal{Q}}(\mathbf{x}, t)\,\mathrm{d}t + \mathbf{L}(\mathbf{x}, t)\,\mathrm{d}\mathbf{B}^{\mathcal{Q}}, \tag{9}$$

where $\mathbf{B}^{\mathcal{Q}}(t)$ denotes an n-dimensional standard Brownian motion. The drift function $\mathbf{f}^{\mathcal{Q}}(\mathbf{x}, t) : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^n$ and dispersion matrix $\mathbf{L}(\mathbf{x}, t) : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^{n\times n}$ are:

$$\begin{aligned}\mathbf{f}(\mathbf{x}, t) &= \boldsymbol{\mathcal{K}}^{\mathcal{Q}}(\boldsymbol{\mu}^{\mathcal{Q}} - \mathbf{x}(t))\\ \mathbf{L}(\mathbf{x}, t) &= \boldsymbol{\Sigma}\,\mathrm{diag}\Big(\sqrt{\boldsymbol{\psi} + \boldsymbol{\Psi}\,\mathbf{x}(t)}\Big).\end{aligned}$$

The parameters $\boldsymbol{\mu}^{\mathcal{Q}}, \boldsymbol{\psi} \in \mathbb{R}^n$ and $\boldsymbol{\mathcal{K}}^{\mathcal{Q}}, \boldsymbol{\Sigma}, \boldsymbol{\Psi} \in \mathbb{R}^{n\times n}$ are time-invariant. With this specification the diffusion matrix of the SDE (9) is $\mathbf{L}(\mathbf{x}, t)\,\mathbf{L}(\mathbf{x}, t)^\top$, which is, in addition to the drift function $\mathbf{f}(\mathbf{x}, t)$, affine in the state $\mathbf{x}(t)$. Note that we denote the probability measure under which the SDE is specified with a superscript on appropriate terms, that is the drift term, its parameters, and the Brownian motion.

Assuming no-arbitrage arguments (Harrison and Kreps, 1979; Harrison and Pliska, 1981) and the short rate process $r(t)$, the time-t price of a zero-coupon bond with maturity at time T, $P(t, T)$, can be solved from the following risk-neutral expectation (see, e.g., Brigo and Mercurio, 2006)

$$P(t, T) = \mathrm{E}^{\mathcal{Q}}\left[e^{-\int_t^T r(\mathbf{x},s)\,\mathrm{d}s} \mid \mathcal{F}_t\right]. \tag{10}$$

Zero-coupon bond prices are a crucial in term structure modeling because the time-t yield curve, consisting of yields $Y(t, T_i) \in \mathbb{R}$ with maturities T_i, is linked to these prices via[3]

$$Y(t, T) = -\frac{1}{T - t}\ln P(t, T). \tag{11}$$

Duffie and Kan (1996) show that with affine term structure models the zero-

[2]The SDE notation of (9) is a standard shorthand way to denote the corresponding Itô stochastic integral equation (see, e.g., Øksendal, 2003).

[3]The zero-coupon bond prices can be stated as $P(t, T) = e^{-Y(t,T)(T-t)}$; for an investment of $P(t, T)$ at time t the rate $Y(t, T)$ *yields* a unit in maturity T.

coupon bond prices solving (10) are exponential affine in the state, that is

$$P(t,T) = e^{a(t)+\mathbf{b}(t)^\top \mathbf{x}(t)}, \tag{12}$$

where $a(t) : \mathbb{R}_+ \to \mathbb{R}$ and $\mathbf{b}(t) : \mathbb{R}_+ \to \mathbb{R}^n$ are solutions to the following system of ordinary differential equations (ODEs)

$$\begin{aligned} \frac{\mathrm{d}a(t)}{\mathrm{d}t} &= \rho_0 - (\boldsymbol{\mathcal{K}}^{\mathcal{Q}}\,\boldsymbol{\mu}^{\mathcal{Q}})^\top \mathbf{b}(t) - \frac{1}{2}\sum_{i=1}^{n} \Big(\boldsymbol{\Sigma}^\top \mathbf{b}(t)\,\mathbf{b}(t)^\top\,\boldsymbol{\Sigma}\Big)_{ii}\,\boldsymbol{\psi}_i \\ \frac{\mathrm{d}\mathbf{b}(t)}{\mathrm{d}t} &= \boldsymbol{\rho}_1 + (\boldsymbol{\mathcal{K}}^{\mathcal{Q}})^\top \mathbf{b}(t) - \frac{1}{2}\sum_{i=1}^{n} \Big(\boldsymbol{\Sigma}^\top \mathbf{b}(t)\,\mathbf{b}(t)^\top\,\boldsymbol{\Sigma}\Big)_{ii}\,(\boldsymbol{\Psi}^\top)_i, \end{aligned} \tag{13}$$

with boundary conditions

$$a(T) = 0, \quad \mathbf{b}(T) = \mathbf{0}.$$

Here $\boldsymbol{\psi}_i$ denotes the ith element of the vector $\boldsymbol{\psi} \in \mathbb{R}^n$, $\boldsymbol{\Psi}_i$ is the ith column of the matrix $\boldsymbol{\Psi}_i \in \mathbb{R}^{n\times n}$ and $(\cdot)_{ii}$ refers to the ith diagonal element of the corresponding matrix.

We conclude by showing how, in principle, the zero-coupon bond prices of ATSMs can be derived — without addressing the necessary technical requirements. This is a worthwhile effort in practice, because the form of ODE system (13) varies ever so slightly with different model specifications put forth in literature. The derivation is based on *Feynman–Kac* formula (see, e.g., Karatzas and Shreve, 1991), which establishes a connection between stochastic processes and partial differential equations (PDEs).

Following Klebaner (2005), the Feynman–Kac formula states that for bounded functions $g(\mathbf{x},t)$ and $r(\mathbf{x},t)$, solution to the expectation

$$g(\mathbf{x},t) = \mathrm{E}^{\mathcal{Q}}\left[e^{-\int_t^T r(\mathbf{x},s)\,\mathrm{d}s} h(\mathbf{x},T)) \mid \mathbf{x}(t) = \mathbf{x}\right]$$

satisfies the following partial differential equation

$$\frac{\partial g(\mathbf{x},t)}{\partial t} + \mathscr{L} g(\mathbf{x},t) - r(\mathbf{x},t)\,g(\mathbf{x},t) = 0, \tag{14}$$

with a boundary condition $g(\mathbf{x}, T) = h(\mathbf{x}, T)$ and an operator $\mathscr{L}$ defined as

$$\mathscr{L} = \mathbf{f}(\mathbf{x}, t)^\top \nabla g(\mathbf{x}, t) + \frac{1}{2}\mathrm{Tr}(\mathbf{L}\,\mathbf{L}^\top \nabla\nabla^\top g(\mathbf{x}, t)).$$

This relationship enables us to derive the result (12) by solving the PDE (14) with a boundary condition $g(\mathbf{x}, T) = 1$ and an interpretation of $r(\mathbf{x}, t)$ as the short rate process, which is defined by the SDE (9). We anticipate a solution of the form $g(\mathbf{x}, t) = e^{a(t)+\mathbf{b}(t)^\top \mathbf{x}(t)}$ and substitute it into (14)

$$e^{a(t)+\mathbf{b}(t)^\top \mathbf{x}(t)}\Bigg[\frac{\mathrm{d}a(t)}{\mathrm{d}t} + \frac{\mathrm{d}\mathbf{b}(t)^\top}{\mathrm{d}t}\mathbf{x}(t) + \mathbf{f}(\mathbf{x}, t)^\top \mathbf{b}(t) + \mathrm{Tr}\Big(\mathbf{L}(\mathbf{x}, t)\,\mathbf{L}(\mathbf{x}, t)^\top \mathbf{b}(t)\,\mathbf{b}(t)^\top\Big) - (\rho_0 + \boldsymbol{\rho}_1^\top \mathbf{x}(t))\Bigg] = 0.$$

Both sides of this equation are then divided with $e^{a(t)+\mathbf{b}(t)^\top \mathbf{x}(t)}$ and the remaining terms are collected and re-arranged so that we have

$$u(t) + v(\mathbf{x}, t) = 0, \tag{15}$$

where

$$\begin{aligned} u(t) &= \frac{\mathrm{d}a(t)}{\mathrm{d}t} + (\boldsymbol{\mathcal{K}}^{\mathcal{Q}}\,\boldsymbol{\mu}^{\mathcal{Q}})^\top \mathbf{b}(t) + \mathrm{Tr}\Big(\boldsymbol{\Sigma}\,\mathrm{diag}(\psi)\,\boldsymbol{\Sigma}^\top \mathbf{b}(t)\,\mathbf{b}(t)^\top\Big) - \rho_0 \\ v(\mathbf{x}, t) &= \frac{\mathrm{d}\mathbf{b}(t)^\top}{\mathrm{d}t}\,\mathbf{x}(t) - (\boldsymbol{\mathcal{K}}^{\mathcal{Q}}\,\mathbf{x}(t))^\top \mathbf{b}(t) \\ &\quad + \mathrm{Tr}\Big(\boldsymbol{\Sigma}\,\mathrm{diag}(\boldsymbol{\Psi}\,\mathbf{x}(t))\,\boldsymbol{\Sigma}^\top \mathbf{b}(t)\,\mathbf{b}(t)^\top\Big) - \boldsymbol{\rho}_1^\top \mathbf{x}(t). \end{aligned}$$

We rewrite $v(\mathbf{x}, t) = \mathbf{x}^\top \tilde{\mathbf{v}}(t)$, where

$$\tilde{\mathbf{v}}(t) = \frac{\mathrm{d}\mathbf{b}(t)}{\mathrm{d}t} - (\boldsymbol{\mathcal{K}}^{\mathcal{Q}})^\top \mathbf{b}(t) + \sum_{i=1}^{n}\Big(\boldsymbol{\Sigma}\,\boldsymbol{\Sigma}^\top \mathbf{b}(t)\,\mathbf{b}(t)^\top\Big)_{ii}(\boldsymbol{\Psi}^\top)_i - \boldsymbol{\rho}_1,$$

and $(\boldsymbol{\Psi}^\top)_i$ is used to denote the ith column of the transposed matrix $\boldsymbol{\Psi}^\top$. Because (15) needs to hold for all $\mathbf{x}(t)$, we require that

$$u(t) = 0 \quad \text{and} \quad \tilde{\mathbf{v}}(t) = \mathbf{0},$$

and the ODE system (13) follows with boundary conditions $a(T) = 0$ and $\mathbf{b}(T) = \mathbf{0}$. This establishes that the zero-coupon prices (12) solve, indeed, the PDE (14). Therefore, they also solve the pricing expectation (10).

3.2 State Dynamics Under the Physical Measure

In order to describe the real-world dynamics of the state, the dynamic model (9) needs to be specified under the physical probability measure $\mathcal{P}$. Without going into details of the Girsanov theorem (see, e.g., Karatzas and Shreve, 1991; Øksendal, 2003), or likelihood ratios, we can change the measure under which the SDE (9) is specified from the risk-neutral $\mathcal{Q}$ to the physical measure $\mathcal{P}$ by substituting the following transformation

$$\mathrm{d}\mathbf{B}^{\mathcal{Q}} = \mathrm{d}\mathbf{B}^{\mathcal{P}} + \boldsymbol{\lambda}(t)\,\mathrm{d}t \tag{16}$$

into the SDE (9) (see, p. 278 in Klebaner, 2005). Here $\boldsymbol{\lambda}(t) : \mathbb{R}_+ \to \mathbb{R}^n$ is the risk premium, also called the market price of risk, whose functional form determines the relationship between state dynamics under the two measures.

Several specifications for the risk premium have been proposed in literature. Dai and Singleton (2000) consider a completely affine risk premium that is nested in the essentially affine specification of Duffee (2002) that, in turn, is developed further by the semi-affine form of Duarte (2004). Most recently, Cheridito *et al.* (2007) introduce an extended market price of risk specification. The issue of risk premium specification is not pursued here, rather, we present the essentially affine market price of risk and refer the reader to the excellent review and comparison of Feldhütter (2008).

The essentially affine risk premium of Duffee (2002) is

$$\boldsymbol{\lambda}(t) = \operatorname{diag}\Big(\sqrt{\mathbf{s}(\mathbf{x},\boldsymbol{\theta})}\Big)\,\boldsymbol{\phi} + \operatorname{diag}\Big(\mathbf{1}_{\mathbf{s}}\,\sqrt{\mathbf{s}^{-1}(\mathbf{x},\boldsymbol{\theta})}\Big)\,\boldsymbol{\Phi}\,\mathbf{x}(t), \tag{17}$$

where $\mathbf{s}(\mathbf{x},t) : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^n$ is the argument of the square root in the dispersion matrix $\mathbf{L}(\mathbf{x},t)$ of the SDE (9), that is

$$\mathbf{s}(\mathbf{x},t) = \boldsymbol{\psi} + \boldsymbol{\Psi}\,\mathbf{x}(t).$$

The risk premium parameter dimensions are $\boldsymbol{\phi} \in \mathbb{R}^n$, $\boldsymbol{\Phi} \in \mathbb{R}^{n\times n}$ and $\mathbf{1}_{\mathbf{s}}$ is an n-dimensional indicator vector defined here as

$$(\mathbf{1}_{\mathbf{s}})_i = \begin{cases} 1, & \text{if } \inf(\mathbf{s}_i(\mathbf{x},t)) > 0 \\ 0, & \text{otherwise} \end{cases}, \quad i = 1,2,\ldots,n.$$

This essentially affine risk premium specification ensure that the dynamic model is affine in the state under the physical measure $\mathcal{P}$ and that the volatility does not explode if some $\mathbf{s}_i(\mathbf{x}, t)$ would approach zero.

We can write the state dynamics under $\mathcal{P}$ measure as follows by substituting the measure change (16), with the risk premium (17), to the SDE (9)

$$\mathrm{d}\mathbf{x}(t) = \mathbf{f}^{\mathcal{P}}(\mathbf{x}, t)\,\mathrm{d}t + \mathbf{L}(\mathbf{x}, t)\,\mathrm{d}\mathbf{B}^{\mathcal{P}}. \tag{18}$$

The dispersion matrix $\mathbf{L}(\mathbf{x}, t)$ is the same under measures $\mathcal{P}$ and $\mathcal{Q}$ while the drift term $\mathbf{f}^{\mathcal{P}}$ is of the same functional form as $\mathbf{f}^{\mathcal{Q}}$, that is,:

$$\begin{aligned}\mathbf{f}^{\mathcal{P}}(\mathbf{x}, t) &= \boldsymbol{\mathcal{K}}^{\mathcal{P}}(\boldsymbol{\mu}^{\mathcal{P}} - \mathbf{x}(t)) \\ \mathbf{L}(\mathbf{x}, t) &= \boldsymbol{\Sigma}\,\mathrm{diag}\Big(\sqrt{\boldsymbol{\psi} + \boldsymbol{\Psi}\,\mathbf{x}(t)}\Big)\,.\end{aligned}$$

The drift term parameters under $\mathcal{P}$ measure, $\boldsymbol{\mu}^{\mathcal{P}} \in \mathbb{R}^n$ and $\mathbf{K}^{\mathcal{P}} \in \mathbb{R}^{n\times n}$, have the following relationship to the risk neutral ones

$$\begin{aligned}\mathbf{K}^{\mathcal{P}} &= \mathbf{K}^{\mathcal{Q}} - \boldsymbol{\Sigma}\begin{bmatrix}\boldsymbol{\phi}_1\,\boldsymbol{\Psi}_1^{\top} \\ \vdots \\ \boldsymbol{\phi}_n\,\boldsymbol{\Psi}_n^{\top}\end{bmatrix} + \boldsymbol{\Sigma}\,\mathrm{diag}(\mathbf{1}_{\mathbf{s}})\,\boldsymbol{\Phi} \\ \mathbf{K}^{\mathcal{P}}\boldsymbol{\mu}^{\mathcal{P}} &= \left(\mathbf{K}^{\mathcal{Q}}\,\boldsymbol{\mu}^{\mathcal{Q}} + \boldsymbol{\Sigma}\,(\boldsymbol{\psi}\circ\boldsymbol{\phi})\right),\end{aligned}$$

where $\boldsymbol{\psi}\circ\boldsymbol{\phi}$ denotes Hadamard product, that is for vectors $(\boldsymbol{\psi}\circ\boldsymbol{\phi})_i = \boldsymbol{\psi}_i\,\boldsymbol{\phi}_i$. The flexible specification of the risk premium $\boldsymbol{\lambda}(t)$ enables one to choose the parameters $\boldsymbol{\mu}^{\mathcal{P}}$ and $\mathbf{K}^{\mathcal{P}}$ independently of $\boldsymbol{\mu}^{\mathcal{Q}}$ and $\mathbf{K}^{\mathcal{Q}}$ in the case of Gaussian state dynamics, i.e. the dispersion matrix is not state dependent ($\boldsymbol{\Psi} = \mathbf{0}$ and $\mathbf{1}_{\mathbf{s}} = \mathbf{I}$). This is straightforward to deduce from the above equations.

Time-evolution of the state under $\mathcal{P}$ measure is determined by the stochastic differential equation (18). The evolution of the state in time can be accomplished by using (18) for simulation. In the case where the dispersion matrix is *not* state dependent, that is $\mathbf{L}(\mathbf{x}, t) = \mathbf{L}$, an exact solution for future state distributions can be obtained, because the SDE is linear in the state. The exact solution for (18) is shown in Jazwinski (1970) to be a Gaussian

$$\mathbf{x}(t) \sim \mathcal{N}(\mathbf{m}(t), \mathbf{C}(t)),$$

where the mean $\mathbf{m}(t)$ and covariance $\mathbf{C}(t)$ at time t are given by

$$\begin{aligned}\mathbf{m}(t) &= \exp(-\boldsymbol{\mathcal{K}}^{\mathcal{P}}\, t)\, \mathbf{m}(0) + (\mathbf{I} - \exp(-\boldsymbol{\mathcal{K}}^{\mathcal{P}}\, t))\boldsymbol{\mu}^{\mathcal{P}} \\ \mathbf{C}(t) &= \exp(-\boldsymbol{\mathcal{K}}^{\mathcal{P}}\, t)\, \mathbf{C}(0)\, \exp(-\boldsymbol{\mathcal{K}}^{\mathcal{P}}\, t)^{\top} \\ &\quad + \int_0^t \exp(-\boldsymbol{\mathcal{K}}^{\mathcal{P}}\, (t-s))\, \mathbf{L}\, \mathbf{L}^{\top}\, \exp(-\boldsymbol{\mathcal{K}}^{\mathcal{P}}\, (t-s))^{\top}\, \mathrm{d}s,\end{aligned} \tag{19}$$

and the state distribution at time $t = 0$ is $\mathbf{x}(0) \sim \mathcal{N}(\mathbf{m}(0), \mathbf{C}(0))$. For pathwise state simulation we can use the above solution by defining a recursion

$$\mathbf{x}(t_{k+1}) = e^{-\boldsymbol{\mathcal{K}}^{\mathcal{P}}\, \Delta t_k}\, \mathbf{x}(t_k) + (\mathbf{I} - e^{-\boldsymbol{\mathcal{K}}^{\mathcal{P}}\, \Delta t_k})\boldsymbol{\mu}^{\mathcal{P}} + \mathbf{q}, \tag{20}$$

where $\Delta t_k = t_{k+1} - t_k$, $\mathbf{q} \sim \mathcal{N}(0, \mathbf{Q}_k)$ and $\mathbf{Q}_k$ is the integral

$$\mathbf{Q}_k = \int_{t_k}^{t_{k+1}} \exp(-\boldsymbol{\mathcal{K}}^{\mathcal{P}}\, (t_{k+1} - s))\, \mathbf{L}\, \mathbf{L}^{\top}\, \exp(-\boldsymbol{\mathcal{K}}^{\mathcal{P}}\, (t_{k+1} - s))^{\top}\, \mathrm{d}s.$$

An exact solution is not generally obtainable for (18) when the dispersion matrix depends on the state; numerical simulation methods need to be used in that case. Kloeden and Platen (1999) provide an extensive review and discussion of different schemes for obtaining numerical solutions to SDEs. In general, the different numerical approximations differ in their order of convergence and whether the produced approximations convergence in the strong or weak sense. The order of convergence represents the rate at which approximation error goes to zero when the step size goes to zero; the larger the order the better. Loosely speaking, strong convergence refers to the solution being a pathwise approximation to the state, where as a weak solution is an approximation for some function of the state at given final time T.

The simplest and most inaccurate method for approximate SDE solutions is the Euler-Maruyama method (Maruyama, 1955), which has a strong convergence order of 0.5. Approximate solutions for (18) can be obtained via

$$\mathbf{X}(t_{k+1}) = \mathbf{X}(t_k) + \mathbf{f}^{\mathcal{P}}(\mathbf{X}, t_k)\, \Delta t_k + \mathbf{L}(\mathbf{X}, t_k) \Delta \mathbf{B}_k^{\mathcal{P}},$$

where $\mathbf{X}(t_k)$ denotes the approximation for $\mathbf{x}(t_k)$, $\Delta t_k = t_{k+1} - t_k$ and $\Delta \mathbf{B}_k^{\mathcal{P}} = \mathbf{B}^{\mathcal{P}}(t_{k+1}) - \mathbf{B}^{\mathcal{P}}(t_k)$.

3.3 Identification and Admissibility

To ensure that there exists a unique strong solution — that does not explode[4] — to the SDEs (9) and (18), the corresponding drift and volatility parameters need to satisfy regularity requirements (Klebaner, 2005; Øksendal, 2003). In applied ATSM literature, model parametrizations for which a unique strong solution exists are sometimes called *admissible*. The question of admissibility is complicated by the fact that generally ATSMs are specified in a non-unique way, that is, two different sets of parameters can generate identical model outputs. To facilitate model identification, Dai and Singleton (2000) present the *canonical form* of ATSMs, which they classify into subsets denoted with $A_M(N)$, where N is the state dimension and M is the number of state variables with state dependent on volatility. Additionally, they provide regularity conditions for the model parameters within the canonical form framework.

We present here a sufficient regularity condition, provided by Duffie and Kan (1996), which ensure the admissibility of an affine term structure model. The condition is overidentifying (demands more than is strictly needed from the parametrization), but it is straightforward compared to the canonical form of ATSMs and the restrictions therein. The regularity condition of Duffie and Kan is as follows:

For all i

(a) For all $\mathbf{x}$ such that $\mathbf{s}_i(\mathbf{x}, t) = 0$,
$$\boldsymbol{\Psi}_i^\top \left(-\boldsymbol{\mathcal{K}}\, \mathbf{x}(t) + \boldsymbol{\mathcal{K}}\, \boldsymbol{\mu}\right) > \frac{1}{2} \boldsymbol{\Psi}_i^\top \, \boldsymbol{\Sigma}\, \boldsymbol{\Sigma}^\top \, \boldsymbol{\Psi}_i.$$

(b) For all j, if $\left(\boldsymbol{\Psi}_i^\top \, \boldsymbol{\Sigma}\right)_j \neq 0$,
$$\text{then } \mathbf{s}_i(\mathbf{x}, t) = k\, \mathbf{s}_j(\mathbf{x}, t), \quad k > 0.$$

Here $\mathbf{s}_i(\mathbf{x}, t)$ denotes the ith component of the vector valued function defined previously as $\mathbf{s}(\mathbf{x}, t) = \boldsymbol{\psi} + \boldsymbol{\Psi}\, \mathbf{x}(t)$ and $\boldsymbol{\Psi}_i$ denotes the ith column of the matrix $\boldsymbol{\Psi}$. In the case of a constant volatility, $\mathbf{L} = \boldsymbol{\Sigma}$, it is enough that the diffusion matrix $\mathbf{L}\mathbf{L}^\top$ is symmetric and positive semi-definite, because $\boldsymbol{\Psi} = \mathbf{0}$ and the above condition holds.

[4] That is, $|\mathbf{x}_i(t)|$ does not tend to infinity.

3.4 Pricing Derivatives

We consider here transformation analysis based pricing of plain vanilla interest rate derivatives with affine term structure models. Key points of the theory as well as more explicit formulas for swaps, zero-coupon bond options and caps (floors) are reviewed. Swaption pricing with ATSMs requires approximative solutions, which we will not present or consider here; different approaches are presented in Singleton and Umantsev (2002) and Schrager and Pelsser (2006). For pricing of exotic interest rate derivatives, see Duffie *et al.* (2000).

Theoretical advances with ATSMs do not yet provide for multi-curve pricing (see, e.g., Bianchetti, 2010) that has evolved to be the market practice after the 2007–2008 financial crisis (Pallavicini and Tarenghi, 2010). Subsequently, the treatment here is built on the premise of single curve pricing, where forward rates and discount factors are computed from the same curve.

Derivative pricing is based on the fundamental assumption of absence of arbitrage opportunities presented in the seminal paper by Black and Scholes (1973). This assumption was later developed into a general mathematical approach of arbitrage-free derivative pricing by Harrison and Kreps (1979) and Harrison and Pliska (1981). Reviewing this theory is beyond the scope of this thesis; we simply assume the existence of a risk-neutral pricing measure Q under which the the time-t price of a derivative with payoffs H_{T_i} at times T_i, $i = 1, 2, \ldots, n$, $T_i > i$, is

$$V(t) = \sum_i \mathrm{E}^{\mathcal{Q}} \left[e^{-\int_t^{T_i} r(\mathbf{x},s)\,\mathrm{d}s} H_i \mid \mathcal{F}_t \right], \tag{21}$$

where $r(\mathbf{x}, s)$ is the short rate process, the filtration $\mathcal{F}_t$ contains the available information up to present time t and the payoffs are assumed to be known at corresponding times without early-exercise possibilities.

In the case of interest rate derivatives, the payoffs H_{T_i} depend on some interest rates (underlyings) that might be fixed prior to the payment date; this is the case with swaps, caps and floors for example. These kind of deferred payoffs can be handled by applying iterated conditioning. Assume a derivative with a payoff at time T that is fixed at time $\tau < T$. The price

of this derivative is given as follows (see p. 42 in Brigo and Mercurio, 2006)

$$\begin{aligned} V(t) &= \mathrm{E}^{\mathcal{Q}}\left[e^{-\int_t^T r(\mathbf{x},s)\,\mathrm{d}s}\, H_\tau \mid \mathcal{F}_t\right] \\ &= \mathrm{E}^{\mathcal{Q}}\left[\mathrm{E}^{\mathcal{Q}}\left[e^{-\int_t^T r(\mathbf{x},s)\,\mathrm{d}s}\, H_\tau \mid \mathcal{F}_\tau\right] \mid \mathcal{F}_t\right] \\ &= \mathrm{E}^{\mathcal{Q}}\left[e^{-\int_t^\tau r(\mathbf{x},s)\,\mathrm{d}s}\, H_\tau\, \mathrm{E}^{\mathcal{Q}}\left[e^{-\int_\tau^T r(\mathbf{x},s)\,\mathrm{d}s} \mid \mathcal{F}_\tau\right] \mid \mathcal{F}_t\right] \\ &= \mathrm{E}^{\mathcal{Q}}\left[e^{-\int_t^\tau r(\mathbf{x},s)\,\mathrm{d}s}\, H_\tau\, P(\tau,T) \mid \mathcal{F}_t\right], \end{aligned}$$

where $P(\tau,T)$ is the time-τ price of a zero-coupon bond.

The breakthrough results of Duffie *et al.* (2000) facilitate pricing of interest rate derivatives with ATSMs through transformation analysis, that is, computation of (21) when the short rate evolution is determined by an ATSM and the underlyings are related to interest rates. In the following, we review two key results presented by these authors.

Firstly, under certain technical regularity conditions (see Duffie *et al.*, 2000) ensuring well-behavedness of the transform $\Gamma(\mathbf{u},\mathbf{x}_t,t,T):\mathbb{C}^n\times\mathbb{R}^n\times\mathbb{R}_+\times\mathbb{R}_+\to\mathbb{C}$, $t\le T$, we have

$$\begin{aligned} \Gamma(\mathbf{u},\mathbf{x}_t,t,T) &= \mathrm{E}^{\mathcal{Q}}\left[e^{-\int_t^T r(\mathbf{x},s)\,\mathrm{d}s}\, e^{\mathbf{u}^\top\mathbf{x}(T)} \mid \mathcal{F}_t\right] \\ &= e^{a(t,T,\mathbf{u})+\mathbf{b}(t,T,\mathbf{u})^\top\mathbf{x}(t)}, \end{aligned} \tag{22}$$

where $a(t,T,\mathbf{u}):\mathbb{R}_+\times\mathbb{R}_+\times\mathbb{C}^n\to\mathbb{C}$ and $\mathbf{b}(t,T,\mathbf{u}):\mathbb{R}_+\times\mathbb{R}_+\times\mathbb{C}^n\to\mathbb{C}^n$ are solutions to the ODE system (13), restated here for convenience,

$$\begin{aligned} \frac{\mathrm{d}a(t)}{\mathrm{d}t} &= \rho_0 - (\boldsymbol{\mathcal{K}}^{\mathcal{Q}}\boldsymbol{\mu}^{\mathcal{Q}})^\top\mathbf{b}(t) - \frac{1}{2}\sum_{i=1}^n\left(\boldsymbol{\Sigma}^\top\mathbf{b}(t)\,\mathbf{b}(t)^\top\boldsymbol{\Sigma}\right)_{ii}\boldsymbol{\psi}_i, \\ \frac{\mathrm{d}\mathbf{b}(t)}{\mathrm{d}t} &= \boldsymbol{\rho}_1 + (\boldsymbol{\mathcal{K}}^{\mathcal{Q}})^\top\mathbf{b}(t) - \frac{1}{2}\sum_{i=1}^n\left(\boldsymbol{\Sigma}^\top\mathbf{b}(t)\,\mathbf{b}(t)^\top\boldsymbol{\Sigma}\right)_{ii}(\boldsymbol{\Psi}^\top)_i, \end{aligned} \tag{23}$$

with boundary conditions

$$a(T)=0,\quad \mathbf{b}(T)=\mathbf{u}.$$

The boundary condition for $\mathbf{b}(t)$ is explicitly indicated in the notation; $a(t,T,\mathbf{u})$ and $\mathbf{b}(t,T,\mathbf{u})$ denote the solutions to the above differential equations evaluated at time t with boundary condition $\mathbf{u}$ for $\mathbf{b}(t)$ at time T.

Note that the time-t zero-coupon bond price $P(t,T)$ equals the transform (22) with zero boundary condition $\mathbf{u} = \mathbf{0}$, that is, we have a relation

$$P(t,T) = \Gamma(\mathbf{0}, \mathbf{x}_t, t, T). \tag{24}$$

Secondly, suppose that $\Gamma(\mathbf{u} + jv\mathbf{q}, \mathbf{x}_0, 0, T)$ is well-behaving for any $v \in \mathbb{R}$ with fixed $\mathbf{x}_0, \mathbf{u}, \mathbf{q} \in \mathbb{R}^n$, and the condition

$$\int_{\mathbb{R}} |\, \Gamma(\mathbf{u} + jv\mathbf{q}, \mathbf{x}_0, 0, T)\, \mathrm{d}v \,| < \infty,$$

where j is the imaginary unit. Then we have a mapping $G(\mathbf{u}, \mathbf{q}, c, \mathbf{x}_0, T) : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}_+$ defined by

$$\begin{aligned} G(\mathbf{u}, \mathbf{q}, c, \mathbf{x}_0, T) &= \mathrm{E}^{\mathcal{Q}} \left[e^{-\int_0^T r(\mathbf{x}, s)\, \mathrm{d}s}\, e^{\mathbf{u}^\top \mathbf{x}(T)}\, 1_{\mathbf{q}^\top \mathbf{x}(T) \le c} \mid \mathcal{F}_0 \right], \\ &= \frac{1}{2}\Gamma(\mathbf{u}, \mathbf{x}_0, 0, T) \\ &\quad - \frac{1}{\pi} \int_0^\infty \frac{1}{v} \mathrm{Im}\big[\Gamma(\mathbf{u} + iv\mathbf{q}, \mathbf{x}_0, 0, T)\, e^{-jcv}\big]\, \mathrm{d}v, \end{aligned} \tag{25}$$

where $\mathrm{Im}(z)$ denotes the imaginary part of $z \in \mathbb{C}$ and the indicator function $1_{\mathbf{q}^\top \mathbf{x}(T) \le c}$ takes value one when the condition of the argument is true and zero otherwise. The mappings (22) and (25) form the foundation for transformation analysis based interest rate derivative pricing with affine term structure models.

3.4.1 Interest Rate Swaps

In a prototypical *interest rate swap* (IRS) two parties agree to exchange interest rate payments in same currency on a notional amount at predefined set of dates. Typically, one counterparty pays a fixed rate on the notional whereas payments of the other counterparty are linked to a floating interest rate that resets according at specified intervals. The different streams of cash flows are referred to as the fixed and floating legs of the swap, respectively.

In a payer-swap (PSW) the fixed rate is paid and floating rate received and *vice versa* in a receiver-swap (RSW). Let us assume for simplicity that the payments of both swap legs occur on same dates which are denoted

with $T_i,\ i = m+1, m+2, \ldots, n$. The floating rate resets then at dates $T_i,\ i = m, m+1, \ldots, n-1$. If we denote the swap notional with N, the fixed rate with K and the year fraction between two consecutive payment dates with $\tau_i = T_i - T_{i-1}$, we have a swap payoff at time T_i of the form

$$H_{T_i} = wN\tau_i(L(T_{i-1}, T_i) - K),$$

where $L(T_{i-1}, T_i)$ is the underlying floating rate at time T_{i-1} for maturity T_i, $w = 1$ for a PSW and $w = -1$ for a RSW.

Clearly, the floating rate $L(\cdot)$ for a given period is not known before the corresponding reset date has occurred. According to the no-arbitrage theory, it is set a priori equal to the prevailing simply compounded *forward rate* for the period (see, e.g., Rebonato, 1996). The present time forward rate $F(T_{i-1}, T_i)$, from time T_{i-1} to T_i, is by definition

$$F(T_{i-1}, T_i) = \frac{1}{T_i - T_{i-1}} \left(\frac{P(0, T_{i-1})}{P(0, T_i)} - 1 \right). \tag{26}$$

The swap present value $V(0, K, T_{m:n})$, i.e. price at time $t = 0$, is then given by the discounted sum of the payoffs (see, e.g., Brigo and Mercurio, 2006)

$$\begin{aligned} V(0, K, T_{m:n}) &= wN \sum_{i=m+1}^{n} P(0, T_i)\tau_i(F(T_{i-1}, T_i) - K) \\ &= P(0, T_m) - P(0, T_n) - \sum_{i=m+1}^{n} P(0, T_i)\tau_i K_f, \end{aligned} \tag{27}$$

where $T_{m:n} = T_m, \ldots, T_n$. This price can be computed with ATSMs by using the formula (24) to obtain the necessary zero-coupon bond prices. The swap value during the contract, say at time $t',\ T_m < t' < T_{m+1}$, is

$$V(t', K, T_{m:n}) = wNP(t', T_{m+1})\tau_i(L(T_m, T_{m+1}) - K) + V(T_{m+1}, K, T_{m+1:n})$$

The fixed rate that renders an interest rate swap fair at present time, that is, it has zero value, is called *swap rate* K_{sw}, and it is given by

$$K_{sw} = \frac{P(0, T_m) - P(0, T_n)}{\sum_{i=m+1}^{n} \tau_i P(0, T_i)}. \tag{28}$$

3.4.2 Zero-Coupon Bond Options

The payoff of a European call option, with a unit notional and strike $K_c \in \mathbb{R}_+$, on a maturity T_m zero-coupon bond, at the option expiration T_p is

$$H_{T_p} = (P(T_p, T_m) - K_c)^+, \tag{29}$$

where $T_p < T_m$ and $(\cdot)^+ = \max(\cdot, 0)$. The time $t = 0$ price of this option is then according to the fundamental pricing formula (21)

$$V_{\text{ZBC}}(K_c, T_p, T_m) = \mathrm{E}^{\mathcal{Q}}\left[e^{-\int_0^{T_p} r(\mathbf{x},s)\,\mathrm{d}s}\,(P(T_p, T_m) - K_c)^+ \mid \mathcal{F}_t\right]. \tag{30}$$

We can write the underlying zero-coupon bond price $P(T_p, T_m)$ in the following form (in the case of ATSMs) by making use of the formula (22),

$$\begin{aligned} P(T_p, T_m) &= \Gamma(\mathbf{0}, \mathbf{x}_{T_p}, T_p, T_m) \\ &= e^{a(T_p,T_m,\mathbf{0}) + \mathbf{b}^\top(T_p,T_m,\mathbf{0})\,\mathbf{x}(T_p)} \\ &= e^{a_{T_p} + \mathbf{b}_{T_p}^\top \mathbf{x}(T_p)}, \end{aligned}$$

where $a_{T_p} = a(T_p, T_m, \mathbf{0})$ and $\mathbf{b}_{T_p} = \mathbf{b}(T_p, T_m, \mathbf{0})$ are the ODE system (23) solutions at time T_p. The above form of $P(T_p, T_m)$ allows us to rewrite the payoff (29) as

$$(P(T_p, T_m) - K_c)^+ = e^{a_{T_p}}\left(e^{\mathbf{b}_{T_p}^\top \mathbf{x}(T_p)} - K_c\, e^{-a_{T_p}}\right) 1_{\mathbf{b}_{T_p}^\top \mathbf{x}(T_p) \geq \ln(K_c) - a_{T_p}},$$

where $1_{\mathbf{b}_{T_p}^\top \mathbf{x}_{T_p} \geq \ln(K_c) - a_{T_p}}$ takes value one when the subscript condition is true and zero otherwise. By substituting this payoff formulation into (30), and denoting $V_{\text{ZBC}} \triangleq V_{\text{ZBC}}(K_c, T_p, T_m)$ and $\tilde{c} = \ln(K_c) - a_{T_p}$, we have

$$\begin{aligned} V_{\text{ZBC}} &= e^{a_{T_p}}\,\mathrm{E}^{\mathcal{Q}}\left[e^{-\int_0^{T_p} r(\mathbf{x},s)\,\mathrm{d}s}\left(e^{\mathbf{b}_{T_p}^\top \mathbf{x}(T_p)} - K_c\, e^{-a_{T_p}}\right) 1_{\mathbf{b}_{T_p}^\top \mathbf{x}(T_p) \geq \tilde{c}} \mid \mathcal{F}_0\right] \\ &= e^{a_{T_p}}\,\mathrm{E}^{\mathcal{Q}}\left[e^{-\int_0^{T_p} r(\mathbf{x},s)\,\mathrm{d}s}\, e^{\mathbf{b}_{T_p}^\top \mathbf{x}(T_p)}\, 1_{\mathbf{b}_{T_p}^\top \mathbf{x}(T_p) \geq \tilde{c}} \mid \mathcal{F}_0\right] \\ &\quad - K_c\,\mathrm{E}^{\mathcal{Q}}\left[e^{-\int_0^{T_p} r(\mathbf{x},s)\,\mathrm{d}s}\, 1_{\mathbf{b}_{T_p}^\top \mathbf{x}(T_p) \geq \tilde{c}} \mid \mathcal{F}_0\right]. \end{aligned}$$

The terms in both of the brackets above are of the form (25), and we have

$$V_{\text{ZBC}} = e^{a_{T_p}}\, G(\mathbf{b}_{T_p}, -\mathbf{b}_{T_p}, -\tilde{c}, \mathbf{x}_0, T_p) - K_c\, G(\mathbf{0}, -\mathbf{b}_{T_p}, -\tilde{c}, \mathbf{x}_0, T_p), \tag{31}$$

where $V_{\text{ZBC}} \triangleq V_{\text{ZBC}}(K_c, T_p, T_m)$ and $\tilde{c} = \ln(K_c) - a_{T_p}$. The payoff of a put option on a zero-coupon bond with similar trade specifications is

$$H_{T_p} = (K_p - P(T_p, T_m))^+,$$

where K_p is the strike rate. Analogously to the call option price, we can derive the price of a European put option on a zero-coupon bond to be

$$V_{\text{ZBP}} = K_p\, G(\mathbf{0}, \mathbf{b}_{T_p}, \tilde{c}, \mathbf{x}_0, T_p) - e^{a_{T_p}}\; G(\mathbf{b}_{T_p}, \mathbf{b}_{T_p}, \tilde{c}, \mathbf{x}_0, T_p), \qquad (32)$$

where $\tilde{c} = \ln(K_p) - a_{T_p}$.

3.4.3 Caps and Floors

An interest rate *cap*, henceforth simply a cap, is a derivative that pays at the end of those periods, in which a floating interest rate has exceeded a pre-specified strike rate K_c fixed on a reset date in the beginning of the corresponding period. A *floor* is a similar derivative paying if the floating interest rate is below the strike rate. In the following, we derive in detail the formula for cap pricing with ATSMs; corresponding formula is also presented in the end for floors.

Let us fix the payment dates $T_{m+1,m+2,\ldots,n}$ of a unit notional cap with associated reset dates $T_{m,m+1,\ldots,n-1}$ and use $L(T_{i-1}, T_i)$ to denote the floating rate at time T_{i-1} with maturity T_i. The cap pays at time T_i

$$\tau_i (L(T_{i-1}, T_i) - K_c)^+,$$

where τ_i is the year fraction between the reset T_{i-1} and payment T_i dates. This one period part of a cap is termed *caplet*; a cap is a series of caplets where the payment date of a caplet is the reset date of the next one, until the cap maturity. To price a cap, it is technically enough to consider caplet pricing because a cap price is a sum of the prices of those caplets that form the cap.

Let us denote the present value of a caplet with $V_{\text{CPL}} \triangleq V_{\text{CPL}}(K_c, T_{i-1}, T_i)$, where T_{i-1} is the option expiry date (reset date), T_i is the maturity (payment date) and $K_c \in \mathbb{R}_+$ is the strike rate. The caplet price can be computed

from the following risk-neutral expectation

$$\begin{aligned} V_{\mathrm{CPL}} &= \mathrm{E}^{\mathcal{Q}}\left[e^{-\int_0^{T_i} r(\mathbf{x},s)\,\mathrm{d}s}\,(L(T_{i-1},T_i)-K_c)^+ \mid \mathcal{F}_0\right] \\ &= \mathrm{E}^{\mathcal{Q}}\left[e^{-\int_0^{T_{i-1}} r(\mathbf{x},s)\,\mathrm{d}s} P(T_{i-1},T_i)\,(L(T_{i-1},T_i)-K_c)^+ \mid \mathcal{F}_0\right], \end{aligned} \tag{33}$$

where we applied the iterated conditioning to obtain the second line formula.

Again we replace $L(T_{i-1},T_i)$ with the corresponding forward rate, that can be written, in addition to the previously presented form (26), as follows

$$F(T_{i-1},T_i) = \frac{1}{\tau}\left(\frac{1}{P(T_{i-1},T_i)} - 1\right).$$

Substituting $L(T_{i-1},T_i) = F(T_{i-1},T_i)$ into the pricing formula (33) yields

$$\begin{aligned} V_{\mathrm{CPL}} &= \mathrm{E}^{\mathcal{Q}}\left[e^{-\int_0^{T_{i-1}} r(\mathbf{x},s)\,\mathrm{d}s}\,P(T_{i-1},T_i)\,\tau(F(T_{i-1},T_i)-K_c)^+ \mid \mathcal{F}_0\right] \\ &= (1+\tau K_c)\mathrm{E}^{\mathcal{Q}}\left[e^{-\int_0^{T_{i-1}} r(\mathbf{x},s)\,\mathrm{d}s}\,\left(\frac{1}{1+\tau K_c} - P(T_{i-1},T_i)\right)^+ \mid \mathcal{F}_0\right], \end{aligned}$$

that is

$$V_{\mathrm{CPL}} = (1+\tau K_c)\,V_{\mathrm{ZBP}}\left(\frac{1}{1+\tau K_c}, T_{i-1}, T_i\right), \tag{34}$$

where $V_{\mathrm{ZBP}}(K_p,T_{i-1},T_i)$ is the present value of a European put option, with a strike $\frac{1}{1+\tau K_c}$ and expiry T_{i-1}, on a maturity T_i zero-coupon bond. By substituting the previously derived formula (32) for computing this put option price, we get the formula for computing caplet price with ATSMs

$$V_{\mathrm{CPL}} = G(\mathbf{0},\mathbf{b}_{T_{i-1}},\tilde{c},\mathbf{x}_0,T_{i-1}) - (1+\tau K_c)e^{a_{T_{i-1}}}\,G(\mathbf{b}_{T_{i-1}},\mathbf{b}_{T_{i-1}},\tilde{c},\mathbf{x}_0,T_{i-1}). \tag{35}$$

Here $a_{T_{i-1}} = a(T_{i-1},T_i,\mathbf{0})$ and $\mathbf{b}_{T_{i-1}} = \mathbf{b}(T_{i-1},T_i,\mathbf{0})$ are the ODE system (23) solutions at time T_{i-1} and $\tilde{c} = \ln(1/(1+\tau K_c)) - a_{T_{i-1}}$.

The price of a cap $V_{\mathrm{CAP}}(t,K_c,T_{m:n})$, at the present time $t=0$, with reset dates $T_{m,m+1,\ldots,n-1}$, payment dates $T_{m+1,m+2,\ldots,n}$ and a strike K_c, is a sum of corresponding caplet prices

$$V_{\mathrm{CAP}}(0,K_c,T_{m:n}) = \sum_{i=m+1}^{n} V_{\mathrm{CPL}}(K_c,T_{i-1},T_i).$$

If the cap price is computed during the contract, for example at time t', $T_m < t' < T_{m+1}$, the already fixed cashflow needs to taken into account. That is, we have

$$V_{\mathrm{CAP}}(t', K_c, T_{m:n}) = \tau_i P(t', T_{m+1})(L(T_m, T_{m+1}) - K_c)^+ + V_{\mathrm{CAP}}(t', K_c, T_{m+1:n}).$$

Analogously to the caplet price formula (34), the time $t = 0$ price of a floorlet is

$$V_{\mathrm{FLL}}(K_p, T_{i-1}, T_i) = (1 + \tau K_p)\, V_{\mathrm{ZBC}}\left(\frac{1}{1+\tau K_p}, T_{i-1}, T_i\right),$$

and the present time floor price is

$$V_{\mathrm{FLR}}(0, K_p, T_{m:n}) = \sum_{i=m+1}^{n} V_{\mathrm{FLL}}(K_p, T_{i-1}, T_i).$$

A cap is termed to be at-the-money (ATM), if its price equals that of a floor with same strike rate and reset/payment dates, that is,

$$V_{\mathrm{CAP}}(0, K_{ATM}, T_{m:n}) = V_{\mathrm{FLR}}(0, K_{ATM}, T_{m:n})$$

The ATM strike rate for caps and floors equals the swap rate (28) of a swap, that has same payment and reset dates as the cap (see, e.g., Brigo and Mercurio, 2006).

3.5 Complete State Space Representation

We recapitulate here the state space model representation of affine term structure models; all the subsequent model components have been presented in more detail in previous sections. The focus here is on the model ensemble and the state space interpretation.

In affine term structure models the dynamic model is defined in continuous time under probability measures $\mathcal{P}$ and $\mathcal{Q}$,

$$\begin{aligned} \mathrm{d}\mathbf{x}(t) &= \mathbf{f}^{\mathcal{Q}}(\mathbf{x}, t)\,\mathrm{d}t + \mathbf{L}(\mathbf{x}, t)\,\mathrm{d}\mathbf{B}^{\mathcal{Q}}, \\ \mathrm{d}\mathbf{x}(t) &= \mathbf{f}^{\mathcal{P}}(\mathbf{x}, t)\,\mathrm{d}t + \mathbf{L}(\mathbf{x}, t)\,\mathrm{d}\mathbf{B}^{\mathcal{P}}, \end{aligned}$$

where under $\mathcal{Q}$ it describes the risk-neutral and under $\mathcal{P}$ the real-world evolution of the state, see the Sections 3.1 and 3.2. Relationship between these two is determined by the market price of risk, see formulas (16) and (17). The measurement model

$$\mathbf{y}(t_k) = \mathbf{h}(\mathbf{x}, t_k) + \mathbf{r},$$

where $\mathbf{r} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ is the Gaussian measurement noise, maps the state into economical quantities, such as yields or derivative prices, or both. Specifically, the function $\mathbf{h}(\mathbf{x}, t_k) : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^k$ is derived from the dynamic model under $\mathcal{Q}$ with the help of short rate specification (8) and no-arbitrage arguments.

Assuming that the state dynamics are rich enough to capture essential underlying features of yield and volatility behavior, joint modeling of interest rates and derivative prices with an ATSM is fairly straightforward; the outputs of the measurement model need to comprise both yields and derivative prices. This can be achieved by specifying the measurement model to incorporate both the linear mapping from state to yields, and the nonlinear mapping from state to derivative prices.

In order to use the model in practice, its parameters need to be estimated. Parameter estimation is often referred to as model calibration in the financial literature. To ensure that the model is admissible and identifiable in the first place, certain restrictions need to be imposed on the parameters, see Section 3.3.

A very important notion regarding model calibration is the following (see, e.g., Aït-Sahalia and Kimmel, 2010): if historical observations are used in the model parameter estimation, the state dynamics are then inferred under the $\mathcal{P}$ measure, although the measurement model is derived from the risk-neutral dynamics and contains, therefore, the parameters under $\mathcal{Q}$-dynamics. Both this notion and the state space modeling framework make it easier to use the Kalman filter in model estimation. This is because Kalman filter allows for simultaneous estimation of dynamic and measurement model parameters.

4 Parameter Estimation in State Space Models

In this chapter, we focus on parameter estimation in state space models (SSMs, Durbin and Koopman, 2012; Barber, 2012) with continuous-discrete dynamics. These models consist of a dynamic model that dictates the time-evolution of the system state observed at discrete times through a measurement model; the continuous-discrete model can be thought of as a limiting case of a discrete time model, where an infinite number of states is added in between the discrete observations (Särkkä, 2006). Both the dynamic and measurement models depend on model parameters that are time-invariant and collected in a vector $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$.

The objective in state space model parameter estimation is to find an estimate for the model parameters $\hat{\boldsymbol{\theta}}$ such that the model outputs are similar to the available observations concerning the system being modeled. Due to the nature of state space models, the parameter estimation problem is tightly coupled with a state estimation (filtering) problem. Hence, we also need to address the state estimation of continuous-discrete state space models in order to answer how to estimate model parameters.

We start by describing a general continuous-discrete state space model, the parameters of which we are keen to estimate. The state dynamics are driven by a continuous-time dynamic model that is an Itô type stochastic differential equation written in a general form of

$$\mathrm{d}\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t, \boldsymbol{\theta})\,\mathrm{d}t + \mathbf{L}(\mathbf{x}, t, \boldsymbol{\theta})\,\mathrm{d}\mathbf{B}(t), \tag{36}$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ is the state vector, $\boldsymbol{\theta}$ contains all model parameters, $\mathbf{f}(\mathbf{x}, t, \boldsymbol{\theta}) : \mathbb{R}^n \times \mathbb{R}_+ \times \Theta \to \mathbb{R}^n$ is a non-random drift function and $\mathbf{L}(\mathbf{x}, t, \boldsymbol{\theta}) : \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}^p \to \mathbb{R}^{n \times n}$ is a non-random dispersion matrix (matrix valued function) and $\mathrm{d}\mathbf{B}(t)$ is n-dimensional standard Brownian motion. The diffusion matrix of (36) is then $\mathbf{L}(\mathbf{x}, t, \boldsymbol{\theta})\,\mathbf{L}(\mathbf{x}, t, \boldsymbol{\theta})^\top$.

Observations of the states are obtained, at discrete times, via measurement model

$$\mathbf{y}_k(\mathbf{x}(t_k), \boldsymbol{\theta}) = \mathbf{h}(\mathbf{x}(t_k), \boldsymbol{\theta}) + \mathbf{r}, \tag{37}$$

where $\mathbf{y}_k \in \mathbb{R}^d$ is an observation at time t_k, $\mathbf{y}_k \triangleq \mathbf{y}(t_k)$, and $\mathbf{r} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ is Gaussian measurement noise with covariance matrix $\mathbf{R} \subseteq \mathbb{R}^{d \times d}$. Because the noise process is additive and Gaussian, we can also write the measurement model in equivalent probabilistic form of

$$p(\mathbf{y}_k \mid \mathbf{x}(t_k), \boldsymbol{\theta}) = \mathcal{N}(\mathbf{y}_k \mid \mathbf{h}(\mathbf{x}(t_k), \boldsymbol{\theta}), \mathbf{R}).$$

Note that in our specification the measurement function $\mathbf{h}(\mathbf{x}(t_k), \boldsymbol{\theta}) : \mathbb{R}^n \times \Theta \to \mathbb{R}^d$ also depends on (some of) the model parameters.

The state dynamics in state space models are assumed to be Markovian. This means that the state sequence produced by the dynamic model incorporates the Markov property, and the observations are conditionally independent. To be more precise, the Markov property of the states means that a future state is only dependent on the present state, not the past states or observations,

$$p(\mathbf{x}(s) \mid \mathbf{x}_{t \leq s}, \mathbf{y}_{1:k \,|\, t_k < s}) = p(\mathbf{x}(s) \mid \mathbf{x}(t)), \quad s \geq t,$$

and the current observation is independent of past observations given the present state (conditional independence)

$$p(\mathbf{y}_k \mid \mathbf{x}_{t \leq t_k}, \mathbf{y}_{1:k-1}) = p(\mathbf{y}_k \mid \mathbf{x}(t_k)),$$

where $\mathbf{y}_{1:k} \triangleq \mathbf{y}(t_1), \mathbf{y}(t_2), \ldots, \mathbf{y}(t_k)$, $\mathbf{x}_{t \leq s} \triangleq \{\mathbf{x}(t) \mid t \leq s\}$ and $\mathbf{y}_{1:k \,|\, t_k < s}$ denotes observations $\mathbf{y}_{1:k}$ where $t_k \leq s$.

The following treatment on how to estimate the parameters of a general continuous-discrete state space model described above is divided into three sections. We start by presenting Bayesian approach to parameter estimation with focus on point estimates, especially on maximum a posteriori (MAP). To cover parameter estimation in full detail, we also need to address a state estimation (filtering) problem. This is done separately for linear and non-linear models.

4.1 Bayesian Point Estimates

In the Bayesian approach to parameter estimation the starting point is to consider parameters as random variables. Prior information regarding parameters $\boldsymbol{\theta}$ is encoded in the *prior probability distribution* $p(\boldsymbol{\theta})$; the prior contains what is known about parameters before seeing observations. After seeing n observations, $\mathbf{y}_{1:n}$, the *posterior probability distribution* $p(\boldsymbol{\theta} \mid \mathbf{y}_{1:n})$ encodes the updated knowledge about parameters. The posterior can be obtained with Bayes' rule

$$p(\boldsymbol{\theta} \mid \mathbf{y}_{1:n}) = \frac{p(\mathbf{y}_{1:n} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{\int p(\mathbf{y}_{1:n} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}}, \tag{38}$$

where the denominator normalizes the product of prior $p(\boldsymbol{\theta})$ and *likelihood* $p(\mathbf{y}_{1:n} \mid \boldsymbol{\theta})$, that is, it ensures that the posterior is a proper probability distribution that integrates to one. Observations affect the posterior only through the likelihood function; basically, the likelihood measures how likely the observations $\mathbf{y}_{1:n}$ are given parameters $\boldsymbol{\theta}$. The less observations there are, the more important the prior.

For practical use in applications, information of the parameters contained in the posterior needs to be shrunk to a point estimate. There are many ways to do this — actually, an infinite number. A decision theoretic (see, e.g., French, 1986; Eisenführ *et al.*, 2010) approach to choosing point estimates is to specify a loss (utility) function $L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ and choose the point estimate $\hat{\boldsymbol{\theta}}$ in a way that minimizes the expected loss, that is

$$\hat{\boldsymbol{\theta}} = \arg\min_{\tilde{\boldsymbol{\theta}}} \int_{\Theta} L(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})\, p(\boldsymbol{\theta} \mid \mathbf{y}_{1:n})\, \mathrm{d}\boldsymbol{\theta}.$$

A popular loss function is the squared loss $L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^\top (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$ which is minimized by choosing the posterior mean

$$\hat{\boldsymbol{\theta}}_{\mathrm{MMSE}} = \int_{\Theta} \boldsymbol{\theta}\, p(\boldsymbol{\theta} \mid \mathbf{y}_{1:n})\, \mathrm{d}\boldsymbol{\theta}$$

as the point estimate. This estimate is commonly referred to as the minimum mean squared error (MMSE) estimate due to the squared loss function it minimizes.

In practice, computation of the full posterior distribution (38) as such is often prohibitive (especially with state space models). To circumvent tractability problems with the full posterior, we work with an unnormalized posterior in parameter estimation

$$\begin{aligned} p(\boldsymbol{\theta} \mid \mathbf{y}_{1:n}) &\propto p(\boldsymbol{\theta})\, p(\mathbf{y}_{1:n} \mid \boldsymbol{\theta}) \\ &= p(\boldsymbol{\theta}) \prod_{k=1}^{n} p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1}, \boldsymbol{\theta}), \end{aligned}$$

where on the second line we have applied repeatedly the rule of conditional probability $p(a, b) = p(a \mid b)\, p(b)$. A negative log-transform of the unnormalized posterior is referred to as the *energy function*

$$E(\boldsymbol{\theta}) = -\log p(\boldsymbol{\theta}) - \sum_{k=1}^{n} \log p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1}, \boldsymbol{\theta}), \tag{39}$$

which can be minimized to obtain the posterior mode, *maximum a posteriori* (MAP), as a point estimate for parameters

$$\hat{\boldsymbol{\theta}}_{\mathrm{MAP}} = \arg\min_{\boldsymbol{\theta}} E(\boldsymbol{\theta}). \tag{40}$$

The posterior mode is same for both normalized and unnormalized posterior, hence unnormalized posterior is used (it makes easier objective function). Note that the MAP estimate coincides with the popular maximum likelihood estimate (MLE) when a uniform prior $p(\boldsymbol{\theta}) = 1$ is specified for parameters.

We need to compute the posterior predictive distribution $p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1}, \boldsymbol{\theta})$ in order to obtain the MAP estimate. In the case of state space models, this is not so straightforward, because an observation $\mathbf{y}_k$ is conditionally independent of the measurement history *given* the state $\mathbf{x}(t_k)$. Therefore, computing $p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1}, \boldsymbol{\theta})$ requires the following integration

$$p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1}, \boldsymbol{\theta}) = \int p(\mathbf{y}_k \mid \mathbf{x}(t_k), \boldsymbol{\theta})\, p(\mathbf{x}(t_k) \mid \mathbf{y}_{1:k-1}, \boldsymbol{\theta})\, \mathrm{d}\mathbf{x}(t_k). \tag{41}$$

The first term in the integral (41) is the measurement model in probabilistic form and the second term is called the predictive distribution. The predictive distribution is at the heart of the state estimation problem, which we need to address to compute the MAP estimate for model parameters.

4.2 Continuous-Discrete State Estimation

In continuous-discrete *state estimation*, the interest lies in inferring the state of a time varying system, modeled with a continuous-discrete state space model, based on noisy measurements. The methodology used for this purpose is called *optimal filtering* (Jazwinski, 1970; Anderson and Moore, 1979). Hence, the state estimation problem is often called (state) filtering problem.

In principle, the required information in continuous-discrete state estimation problem is contained in the posterior distribution of the continuous state given the discrete observations

$$p(\{\mathbf{x}(t) \mid t_1 \leq t \leq t_n\} \mid \mathbf{y}_{1:n}).$$

This probability distribution could be computed with the help of Bayes' rule. In practice, however, computing the full posterior is very burdensome and not feasible in real-time applications where it needs to be re-computed every time a new observation is obtained. Therefore, the objective in continuous-discrete state estimation is to compute the *filtering distributions*

$$p(\mathbf{x}(t) \mid \mathbf{y}_{1:k}), \quad t \in [t_k, t_{k+1}), \tag{42}$$

for all $k = 1, 2, \ldots, n$. These distributions are continuous between observations, $t \in (t_k, t_{k+1})$, but have discontinuities at the observation times t_k.

The filtering distributions (42) can be computed recursively in two distinct computational steps:

1. **Prediction step**: Obtain the *predictive distribution* at time $t > t_{k-1}$, $p(\mathbf{x}(t) \mid \mathbf{y}_{1:k-1}) \triangleq p(\mathbf{x}, t)$, by solving Kolmogorov's forward equation

$$\begin{aligned} \frac{\partial\, p(\mathbf{x}, t)}{\partial t} = &- \sum_i \frac{\partial}{\partial x_i} f_i(\mathbf{x}, t)\, p(\mathbf{x}, t) \\ &+ \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} [\mathbf{L}(\mathbf{x}, t)\, \mathbf{L}^\top(\mathbf{x}, t)]_{ij}\, p(\mathbf{x}, t), \end{aligned} \tag{43}$$

supplemented with a boundary condition (see, e.g. Jazwinski, 1970)

$$p(\mathbf{x}, t_{k-1}) = p(\mathbf{x}(t_{k-1}) \mid \mathbf{y}_{1:k-1}).$$

2. **Update step**: Obtain the time t_k marginal state distribution by applying the Bayes' rule to update the predictive distribution to incorporate information provided by the current observation $\mathbf{y}_k$

$$p(\mathbf{x}(t_k) \mid \mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k \mid \mathbf{x}(t_k))\, p(\mathbf{x}(t_k) \mid \mathbf{y}_{1:k-1})}{\int p(\mathbf{y}_k \mid \mathbf{x}(t_k))\, p(\mathbf{x}(t_k) \mid \mathbf{y}_{1:k-1})\, \mathrm{d}\mathbf{x}(t_k)}. \tag{44}$$

In practice, equations both in the predictive and update steps are developed further rather than used directly; the partial differential equation (in predictive step) is in many cases infeasible to solve as such, and the use of Bayes' rule can be simplified further by taking into account the method and model specifics. Therefore, the above recursion underlies the more specifically developed filtering recursions.

In the following two subsections we present the state filtering problem in context of linear and nonlinear models. This allows us the develop the filtering recursion further and present the computational steps used in practice.

4.2.1 Linear Models

Here we consider the state estimation of a linear continuous-discrete state space model

$$\begin{aligned} \mathrm{d}\mathbf{x}(t) &= (\mathbf{F}\,\mathbf{x}(t) + \mathbf{u})\,\mathrm{d}t + \mathbf{L}\,\mathrm{d}\mathbf{B}(t) \\ \mathbf{y}_k &= \mathbf{H}\,\mathbf{x}(t) + \mathbf{r}, \end{aligned} \tag{45}$$

where $\mathbf{r} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$. It is assumed that the model parameters are both known and not time-dependent and the dispersion matrix is not allowed to be state-dependent (this constraint is relaxed with nonlinear models). In the following, we present a closed form solution to the filtering problem of the state space model (45). This solution is known as the *continuous-discrete Kalman filter* (see, e.g., Jazwinski, 1970).

Assuming $\mathbf{x}(0) \sim \mathcal{N}(\mathbf{m}(0), \mathbf{P}(0))$, the predictive distribution for the linear

system (45) is Gaussian (see, e.g., Jazwinski, 1970),

$$p(\mathbf{x}(t) \mid \mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{x}(t) \mid \mathbf{m}(t), \mathbf{P}(t)), \quad t > t_{k-1},$$

where the mean and covariance are determined by the differential equations

$$\frac{\mathrm{d}\mathbf{m}(t)}{\mathrm{d}t} = \mathbf{F}\,\mathbf{m}(t) + \mathbf{u} \tag{46}$$

$$\frac{\mathrm{d}\mathbf{P}(t)}{\mathrm{d}t} = \mathbf{F}\,\mathbf{P}(t) + \mathbf{P}(t)\,\mathbf{F}^\top + \mathbf{L}\,\mathbf{L}^\top, \tag{47}$$

with initial conditions $\mathbf{m}(t_{k-1}) = \mathbf{m}_{k-1|k-1}$ and $\mathbf{P}(t_{k-1}) = \mathbf{P}_{k-1|k-1}$. These differential equations are derived from Kolmogorov's forward equation (43), and the initial mean and covariance, $\mathbf{m}_{k-1|k-1}$ and $\mathbf{P}_{k-1|k-1}$, are obtained on the previous (time t_{k-1}) update step. Note that mean and covariance characterize completely a Gaussian distribution.

Let us denote the time t_k predictive mean and covariance with $\mathbf{m}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$, respectively. We can solve these analytically from the differential equations to obtain (46) and (47)

$$\begin{aligned}
\mathbf{m}_{k|k-1} &= \exp(\mathbf{F}\,(t_k - t_{k-1}))\,\mathbf{m}_{k-1|k-1} + [-\mathbf{I} + \exp(\mathbf{F}\,(t_k - t_{k-1}))]\,\mathbf{F}^{-1}\mathbf{u} \\
\mathbf{P}_{k|k-1} &= \exp(\mathbf{F}\,(t_k - t_{k-1}))\,\mathbf{P}_{k-1|k-1}\,\exp(\mathbf{F}\,(t_k - t_{k-1}))^\top \\
&\quad + \int_{t_{k-1}}^{t_k} \exp(\mathbf{F}\,(t_k - s))\,\mathbf{L}\,\mathbf{L}^\top\,\exp(\mathbf{F}\,(t_k - s))^\top\,\mathrm{d}s.
\end{aligned} \tag{48}$$

Another, computationally efficient, way of solving the covariance in (47) is by matrix fractions (Grewal and Andrews, 2001; Särkkä, 2006). That is, if we let $\mathbf{P}(t) = \mathbf{C}(t)\,\mathbf{D}^{-1}(t)$ and define a system

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \mathbf{C}(t) \\ \mathbf{D}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{F} & \mathbf{L}\,\mathbf{L}^\top \\ \mathbf{0} & -\mathbf{F}^\top \end{bmatrix} \begin{bmatrix} \mathbf{C}(t) \\ \mathbf{D}(t) \end{bmatrix}$$

with initial conditions $\mathbf{C}(t_{k-1}) = \mathbf{P}_{k-1|k-1}$ and $\mathbf{D}(t_{k-1}) = \mathbf{I}$, then the solution of this system,

$$\begin{bmatrix} \mathbf{C}(t_k) \\ \mathbf{D}(t_k) \end{bmatrix} = \exp\left(\begin{bmatrix} \mathbf{F} & \mathbf{L}\,\mathbf{L}^\top \\ \mathbf{0} & -\mathbf{F}^\top \end{bmatrix} \Delta t_k\right) \begin{bmatrix} \mathbf{P}_{k-1|k-1} \\ \mathbf{I} \end{bmatrix}, \quad \Delta t_k = t_k - t_{k-1}$$

also solves (47) and we have $\mathbf{P}_{k|k-1} = \mathbf{C}(t_k)\,\mathbf{D}^{-1}(t_k)$.

The update step in linear continuous-discrete filtering coincides with the linear discrete time filter, because in both cases the measurement model is specified in discrete time. Therefore, we simply borrow the well-known equations for computing the marginal state distribution (see, e.g., Grewal and Andrews, 2001)

$$\begin{aligned} \mathbf{S}_k &= \mathbf{H}\,\mathbf{P}_{k|k-1}\,\mathbf{H}^\top + \mathbf{R} \\ \mathbf{K}_k &= \mathbf{P}_{k|k-1}\,\mathbf{H}^\top\,\mathbf{S}_k^{-1} \\ \mathbf{m}_{k|k} &= \mathbf{m}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}\,\mathbf{m}_{k|k-1}) \\ \mathbf{P}_{k|k} &= \mathbf{P}_{k|k-1} - \mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^\top, \end{aligned} \tag{49}$$

that is a Gaussian

$$p(\mathbf{x}(t_k) \mid \mathbf{y}_{1:k}) = \mathcal{N}(\mathbf{x}(t_k) \mid \mathbf{m}_{k|k}, \mathbf{P}_{k|k}).$$

For an accessible derivation of the above equations, see Särkkä (2006).

In parameter estimation of state space models, the likelihood of observations (given parameters), $p(\mathbf{y}_{1:n} \mid \boldsymbol{\theta})$, is needed. This can be computed by factorizing the likelihood $p(\mathbf{y}_{1:n} \mid \boldsymbol{\theta}) = \prod_{k=1}^{n} p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1}, \boldsymbol{\theta})$ and by solving the posterior predictive distributions $p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1}, \boldsymbol{\theta})$ while conducting the state filtering recursion. In the case of linear continuous-discrete problem, the distributions $p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1}, \boldsymbol{\theta})$ are Gaussian:

$$\begin{aligned} p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1}) &= \int p(\mathbf{y}_k \mid \mathbf{x}(t_k))\, p(\mathbf{x}(t_k) \mid \mathbf{y}_{1:k-1})\, \mathrm{d}\mathbf{x}(t_k) \\ &= \int \mathcal{N}(\mathbf{y}_k \mid \mathbf{H}\,\mathbf{x}(t_k), \mathbf{R})\, \mathcal{N}(\mathbf{x}(t_k) \mid \mathbf{m}_{k|k-1}, \mathbf{P}_{k|k-1})\, \mathrm{d}\mathbf{x}(t_k) \\ &= \mathcal{N}(\mathbf{y}_k \mid \mathbf{H}\,\mathbf{m}_{k|k-1}, \mathbf{S}_k). \end{aligned}$$

We now summarize this continuous-discrete Kalman filter recursion for linear state filtering. First, initialize $\mathbf{x}(0) \sim \mathcal{N}(\mathbf{m}(0), \mathbf{P}(0))$. Then run the following steps on all observations $1, 2, \ldots, n$.

1. **Prediction step**: Obtain the predictive distribution

$$p(\mathbf{x}(t_k) \mid \mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{x}(t_k) \mid \mathbf{m}_{k|k-1}, \mathbf{P}_{k|k-1})$$

by solving the predictive mean $\mathbf{m}_{k|k-1}$ and covariance $\mathbf{P}_{k|k-1}$ from the differential equations (46) and (47).

2. **Update step**: Obtain the time-t_k marginal state distribution

$$p(\mathbf{x}(t_k) \mid \mathbf{y}_{1:k}) = \mathcal{N}(\mathbf{x}(t_k) \mid \mathbf{m}_{k|k}, \mathbf{P}_{k|k})$$

by computing the mean $\mathbf{m}_{k|k}$ and covariance $\mathbf{P}_{k|k}$ from equations (49).

For computing the likelihood via factorization, the following additional step should be added to the filtering recursion, so that the posterior predictive distributions are also obtained.

3. **Likelihood step**: Obtain the posterior predictive distribution

$$p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{y}_k \mid \mathbf{H}\,\mathbf{m}_{k|k-1}, \mathbf{S}_k)$$

by utilizing $\mathbf{H}\,\mathbf{m}_{k|k-1}$ and $\mathbf{S}_k$ that are computed on the update step.

4.2.2 Nonlinear Models

We consider here the state estimation of a continuous-discrete state space model

$$\begin{aligned} \mathrm{d}\mathbf{x}(t) &= \mathbf{f}(\mathbf{x}, t)\,\mathrm{d}t + \mathbf{L}(\mathbf{x}, t)\,\mathrm{d}\mathbf{B}(t) \\ \mathbf{y}_k &= \mathbf{h}(\mathbf{x}, t) + \mathbf{r}, \end{aligned} \tag{50}$$

$\mathbf{r} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$. Model (50) is nonlinear in the state in both the dynamic and measurement models. We assume that the model parameters are both known and not time-dependent. Note that the dispersion matrix is allowed to be state-dependent, contrary to the treatment with linear models.

We approach the nonlinear filtering problem from a *Gaussian filtering* point of view (see, e.g., Ito and Xiong, 2000). In Gaussian filtering the principle idea is to employ a Gaussian approximation for the filtering distributions

$$p(\mathbf{x}(t) \mid \mathbf{y}_{1:k}) \approx \mathcal{N}(\mathbf{x}(t) \mid \mathbf{m}(t), \mathbf{P}(t)), \quad t \in [t_k, t_{k+1}). \tag{51}$$

As a consequence, several Gaussian type integrals

$$\int \mathbf{g}(\mathbf{x}, t)\, \mathcal{N}(\mathbf{x} \mid \mathbf{m}, \mathbf{P})\, \mathrm{d}\mathbf{x}$$

need to be computed on the prediction and update steps. This can be done with the integration method of choice — a clear advantage of the Gaussian filtering approach. In this thesis we rely on the 3rd degree spherical-radial cubature rule (Wu *et al.*, 2006; Arasaratnam, 2009) to tackle the multidimensional Gaussian integrals of interest. The filtering method relying on this integration scheme falls under the category of sigma point methods and is called the *cubature Kalman filter* (CKF, Arasaratnam and Haykin, 2009).

In the 3rd degree spherical-radial cubature integration an n-dimensional, i.e. $\mathbf{x} \in \mathbb{R}^n$, Gaussian integral is approximated with a weighted sum

$$\int \mathbf{g}(\mathbf{x}, t)\, \mathcal{N}(\mathbf{x} \mid \mathbf{m}, \mathbf{P})\, \mathrm{d}\mathbf{x} \approx \sum_i W^{(i)}\, \mathbf{g}(\mathbf{x}^{(i)}, t). \tag{52}$$

The sigma (cubature) points $\mathbf{x}^{(i)}$ are

$$\mathbf{x}^{(i)} = \mathbf{m} + \sqrt{\mathbf{P}}\, \boldsymbol{\xi}_i,$$

where $\sqrt{\mathbf{P}}\, \sqrt{\mathbf{P}}^{\top} = \mathbf{P}$. The weights in the summation are simply $W^{(i)} = 1/(2n)$ and the vectors $\boldsymbol{\xi}_i$ are of the form

$$\boldsymbol{\xi}_i = \begin{cases} \sqrt{n}\, \mathbf{e}_i & , i = 1, 2, \ldots, n \\ -\sqrt{n}\, \mathbf{e}_{i-n} & , i = n+1, n+2, \ldots, 2n \end{cases} \tag{53}$$

where $\mathbf{e}_i$ is a Cartesian unit vector to the direction of coordinate axis i.

Having presented the means to compute Gaussian integrals, we proceed to present the cubature Kalman filter applied to a continuous-discrete nonlinear filtering problem. The treatment here follows fairly closely that of Särkkä and Solin (2012).

The Gaussian approximation to the filtering distributions (51) is characterized by its mean and covariance (like any Gaussian distribution). Kolmogorov's forward equation (43) gives rise to the following differential equa-

tions which determine the mean and covariance of the predictive distribution

$$\begin{aligned}\frac{\mathrm{d}\mathbf{m}(t)}{\mathrm{d}t} &= \mathrm{E}[\mathbf{f}(\mathbf{x},t)] \\ \frac{\mathrm{d}\mathbf{P}(t)}{\mathrm{d}t} &= \mathrm{E}[\mathbf{f}(\mathbf{x},t)\,(\mathbf{x}(t)-\mathbf{m}(t))^\top] + \mathrm{E}[(\mathbf{x}(t)-\mathbf{m}(t))\,\mathbf{f}^\top(\mathbf{x},t)] \\ &\quad + \mathrm{E}[\mathbf{L}(\mathbf{x},t)\,\mathbf{L}^\top(\mathbf{x},t)].\end{aligned} \tag{54}$$

This system is solved from time t_{k-1}, with initial conditions $\mathbf{m}(t_{k-1}) = \mathbf{m}_{k-1|k-1}$ and $\mathbf{P}(t_{k-1}) = \mathbf{P}_{k-1|k-1}$, where $\mathbf{m}_{k-1|k-1}$ and $\mathbf{P}_{k-1|k-1}$ are obtained on the previous update step.

The expectations in the differential equations (54) are with respect to the true distribution of $\mathbf{x}(t)$ and therefore, cannot be solved without solving the actual forward equation (43). This is circumvented in Gaussian filtering by taking the expectations with respect to the Gaussian approximation of the state distribution, that is, we take the expectations in (54) to be

$$\begin{aligned}\mathrm{E}[\mathbf{f}(\mathbf{x},t)] &= \int \mathbf{f}(\mathbf{x},t)\,\mathcal{N}(\mathbf{x}\mid\mathbf{m},\mathbf{P})\,\mathrm{d}\mathbf{x} \\ \mathrm{E}[\mathbf{f}(\mathbf{x},t)\,(\mathbf{x}-\mathbf{m})^\top] &= \int \mathbf{f}(\mathbf{x},t)\,(\mathbf{x}-\mathbf{m})^\top\,\mathcal{N}(\mathbf{x}\mid\mathbf{m},\mathbf{P})\,\mathrm{d}\mathbf{x} \\ \mathrm{E}[\mathbf{L}(\mathbf{x},t)\,\mathbf{L}^\top(\mathbf{x},t)] &= \int \mathbf{L}(\mathbf{x},t)\,\mathbf{L}^\top(\mathbf{x},t)\,\mathcal{N}(\mathbf{x}\mid\mathbf{m},\mathbf{P})\,\mathrm{d}\mathbf{x}.\end{aligned} \tag{55}$$

Note that here we have temporarily omitted the time dependencies to ease reading, that is, we have denoted $\mathbf{x} = \mathbf{x}(t)$, $\mathbf{m} = \mathbf{m}(t)$ and $\mathbf{P} = \mathbf{P}(t)$.

We can apply the cubature integration scheme (52) to compute the Gaussian integrals (55). Let us denote the prediction step sigma points with

$$\mathbf{x}_p^{(i)} = \mathbf{m}(t) + \sqrt{\mathbf{P}(t)}\,\boldsymbol{\xi}_i,$$

and we can write the differential equations (54) as

$$\begin{aligned}
\frac{\mathrm{d}\mathbf{m}(t)}{\mathrm{d}t} &= \frac{1}{2n}\sum_{i=1}^{2n}\mathbf{f}(\mathbf{m}(t)+\sqrt{\mathbf{P}(t)}\,\boldsymbol{\xi}_i, t)\\
\frac{\mathrm{d}\mathbf{P}(t)}{\mathrm{d}t} &= \frac{1}{2n}\sum_{i=1}^{2n}\mathbf{f}(\mathbf{m}(t)+\sqrt{\mathbf{P}(t)}\,\boldsymbol{\xi}_i, t)\,(\sqrt{\mathbf{P}(t)}\,\boldsymbol{\xi}_i)^\top\\
&\quad+\frac{1}{2n}\sum_{i=1}^{2n}\sqrt{\mathbf{P}(t)}\,\boldsymbol{\xi}_i\,\mathbf{f}^\top(\mathbf{m}(t)+\sqrt{\mathbf{P}(t)}\,\boldsymbol{\xi}_i, t)\\
&\quad+\frac{1}{2n}\sum_{i=1}^{2n}\mathbf{L}(\mathbf{m}(t)+\sqrt{\mathbf{P}(t)}\,\boldsymbol{\xi}_i, t)\,\mathbf{L}^\top(\mathbf{m}(t)+\sqrt{\mathbf{P}(t)}\,\boldsymbol{\xi}_i, t)
\end{aligned} \tag{56}$$

with the initial conditions $\mathbf{m}(t_{k-1}) = \mathbf{m}_{k-1|k-1}$ and $\mathbf{P}(t_{k-1}) = \mathbf{P}_{k-1|k-1}$. From this system we can obtain the predictive mean and covariance at time t_k, which we denote with $\mathbf{m}(t_k) \triangleq \mathbf{m}_{k|k-1}$ and $\mathbf{P}(t_k) \triangleq \mathbf{P}_{k|k-1}$.

Assume that we have obtained $\mathbf{m}(t_k) = \mathbf{m}_{k|k-1}$ and $\mathbf{P}(t_k) = \mathbf{P}_{k|k-1}$ from (56). Next we update the predictive distribution to incorporate a new observations at time t_k. The goal is to form a Gaussian approximation for $p(\mathbf{x}(t_k) \mid \mathbf{y}_{1:k})$. For this end, the following integrals need to be computed

$$\begin{aligned}
\boldsymbol{\mu}_k &= \int \mathbf{h}(\mathbf{x}, t_k)\,\mathcal{N}(\mathbf{x}(t_k) \mid \mathbf{m}_{k|k-1}, \mathbf{P}_{k|k-1})\,\mathrm{d}\mathbf{x}(t_k)\\
\mathbf{S}_k &= \int (\mathbf{h}(\mathbf{x}, t_k) - \boldsymbol{\mu}_k)\,(\mathbf{h}(\mathbf{x}, t_k) - \boldsymbol{\mu}_k)^\top\,\mathcal{N}(\mathbf{x}(t_k) \mid \mathbf{m}_{k|k-1}, \mathbf{P}_{k|k-1})\,\mathrm{d}\mathbf{x}(t_k)\\
\mathbf{C}_k &= \int (\mathbf{x}(t_k) - \mathbf{m}_{k|k-1})\,(\mathbf{h}(\mathbf{x}, t_k) - \boldsymbol{\mu}_k)^\top\,\mathcal{N}(\mathbf{x}(t_k) \mid \mathbf{m}_{k|k-1}, \mathbf{P}_{k|k-1})\,\mathrm{d}\mathbf{x}(t_k)
\end{aligned} \tag{57}$$

so that the update step can be completed with equations familiar from the linear case

$$\begin{aligned}
\mathbf{K}_k &= \mathbf{C}_k\,\mathbf{S}_k^{-1}\\
\mathbf{m}_{k|k} &= \mathbf{m}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \boldsymbol{\mu}_k)\\
\mathbf{P}_{k|k} &= \mathbf{P}_{k|k-1} - \mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^\top,
\end{aligned} \tag{58}$$

where $\mathbf{K}_k$ is the Kalman gain, $\mathbf{m}_{k|k}$ and $\mathbf{P}_{k|k}$ are the filtering distribution mean and covariance.

We apply again the integration scheme (52) to compute the integrals (57). This time however, the sigma points need to be defined with respect to the predictive distribution — see the predictive mean and covariance in the

integrand. We define the update step sigma points as

$$\mathbf{x}_u^{(i)} = \mathbf{m}_{k|k-1} + \sqrt{\mathbf{P}_{k|k-1}}\,\boldsymbol{\xi}_i, \tag{59}$$

where $\xi_\mathbf{i}$ is given in (53). Applying the spherical-radial integration rule to the Gaussian integrals (57) yields

$$\begin{aligned}
\boldsymbol{\mu}_k &= \frac{1}{2n}\sum_{i=1}^{2n} \mathbf{h}(\mathbf{x}_u^{(i)}, t_k) \\
\mathbf{S}_k &= \frac{1}{2n}\sum_{i=1}^{2n} (\mathbf{h}(\mathbf{x}_u^{(i)}, t_k) - \boldsymbol{\mu}_k)\,(\mathbf{h}(\mathbf{x}_u^{(i)}, t_k) - \boldsymbol{\mu}_k)^\top \\
\mathbf{C}_k &= \frac{1}{2n}\sum_{i=1}^{2n} (\mathbf{x}_u^{(i)} - \mathbf{m}_{k|k-1})\,(\mathbf{h}(\mathbf{x}_u^{(i)}, t_k) - \boldsymbol{\mu}_k)^\top.
\end{aligned} \tag{60}$$

The mean $\mathbf{m}_{k|k}$ and covariance $\mathbf{P}_{k|k}$ can be obtained by first computing the quantities in (60) and then using (58) to complete the update step.

In parameter estimation we also need to compute (repeatedly) the energy function (39), which consists of the parameter prior and the posterior predictive distributions $p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1})$. In the cubature Kalman filter, the posterior predictive distribution is given by

$$\begin{aligned}
p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1}) &= \int p(\mathbf{y}_k \mid \mathbf{x}(t_k))\, p(\mathbf{x}(t_k) \mid \mathbf{y}_{1:k-1})\, \mathrm{d}\mathbf{x}(t_k) \\
&= \int \mathcal{N}(\mathbf{y}_k \mid \mathbf{h}(\mathbf{x}, t_k), \mathbf{R})\, \mathcal{N}(\mathbf{x}(t_k) \mid \mathbf{m}_{k|k-1}, \mathbf{P}_{k|k-1})\, \mathrm{d}\mathbf{x}(t_k) \\
&= \mathcal{N}(\mathbf{y}_k \mid \boldsymbol{\mu}_k, \mathbf{S}_k).
\end{aligned}$$

We summarize here the presented continuous-discrete cubature Kalman filter. After initializing $\mathbf{x}(0) \sim \mathcal{N}(\mathbf{m}(0), \mathbf{P}(0))$ the following steps are run on all observations $1, \ldots, n$.

1. **Prediction step**: Obtain the approximative predictive distribution

$$p(\mathbf{x}(t_k) \mid \mathbf{y}_{1:k-1}) \approx \mathcal{N}(\mathbf{x}(t_k) \mid \mathbf{m}_{k|k-1}, \mathbf{P}_{k|k-1})$$

   by solving the predictive mean $\mathbf{m}_{k|k-1}$ and covariance $\mathbf{P}_{k|k-1}$ from differential equations (56).

2. **Update step**: Obtain the approximative filtering distribution

$$p(\mathbf{x}(t_k) \mid \mathbf{y}_{1:k}) \approx \mathcal{N}(\mathbf{x}(t_k) \mid \mathbf{m}_{k|k}, \mathbf{P}_{k|k})$$

by first computing the quantities in (60) and then solving the mean $\mathbf{m}_{k|k}$ and covariance $\mathbf{P}_{k|k}$ from equations (60) with sigma points given in (59).

If the likelihood function $p(\mathbf{y}_{1:n} \mid \boldsymbol{\theta})$ is of interest, the following additional step should be added to the state filtering iteration, so that the likelihood can be computed via factorization $p(\mathbf{y}_{1:n} \mid \boldsymbol{\theta}) = \prod_{k=1}^{n} p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1}, \boldsymbol{\theta})$.

3 **Likelihood step**: Obtain the approximative distribution

$$p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1}, \boldsymbol{\theta}) \approx \mathcal{N}(\mathbf{y}_k \mid \boldsymbol{\mu}_k, \mathbf{S}_k),$$

where $\boldsymbol{\mu}_k$ and $\mathbf{S}_k$ are computed on the update step.

5 Case Studies

In this chapter we review a particular affine term structure model, the affine arbitrage free Nelson–Siegel (AFNS) model of Christensen *et al.* (2011), and apply it to exposure modeling within the framework presented in the Section 2.3. The model provides an interesting case for studying exposure modeling with affine term structure models for a number of reasons. First, being an ATSM, the model can be applied to both risk factor simulation and derivative pricing obviating the need for having separate models for these tasks. Second, Christensen *et al.* (2011) report encouraging results in yield curve modeling with the AFNS model. This implies that the model is a promising candidate for exposure modeling of interest rate derivatives, which all depend on the yield curve, in one way or another. Third, the AFNS model features constant volatility which greatly facilitates the model use from computational point of view, because analytical formula for caplet prices, in addition to zero-coupon bond prices, can be derived. On the other hand, the constant volatility feature is a clear restriction on the model and it is of great interest to study how well the model copes with pricing of derivatives that are also dependent on the volatility related to the yield curve

The AFNS model is applied to exposure modeling in two different case studies. In the first one, the exposure related to a standard interest rate swap is modeled. Swaps are priced straight from the yield curve and hence, we are concerned here basically with yield curve modeling and simulation. The model calibration problem in this case is linear and therefore we apply the linear continuous-discrete Kalman filter, presented in the Section 4.2.1, to parameter estimation. In the second case study, we model the exposure of a derivative portfolio that consists of an interest rate cap and swap. In contrast to the first case study, this model calibration problem is nonlinear, because cap prices are used in model calibration and the measurement model for these is nonlinear. As a consequence, we make use of the continuous-discrete cubature Kalman filter, presented in the Section 4.2.2, in model parameter estimation.

5.1 The AFNS Model

The affine arbitrage free Nelson–Siegel (AFNS) model of Christensen *et al.* (2011) is founded on the idea of transforming the static Nelson–Siegel yield curve model (Nelson and Siegel, 1987) into a dynamic one and forcing the resulting model to be arbitrage free. Following Diebold and Li (2006), the static model for zero-coupon bond yields $Y(\tau)$ with time to maturity τ,

$$Y(\tau) = \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 \left(\frac{1 - e^{\lambda\tau}}{\lambda\tau} \right) + \boldsymbol{\beta}_3 \left(\frac{1 - e^{\lambda\tau}}{\lambda\tau} - e^{\lambda\tau} \right), \quad \lambda, \boldsymbol{\beta}_i \in \mathbb{R}, \tag{61}$$

is converted into a dynamic one by letting the coefficients $\boldsymbol{\beta}_i$ to vary in time — they can be then interpreted as the time-varying level, slope and curvature of the yield curve (Christensen *et al.*, 2011). In the AFNS model, these time-varying coefficients, $\boldsymbol{\beta}_i(t)$, are taken to be state variables $\mathbf{x}_i(t)$, and the model is parametrized so that the measurement model for yields is of the static Nelson-Siegel form (61) supplemented with a "*yield-adjustment term*".

In the AFNS model, the state is three-dimensional $\mathbf{x}(t) \in \mathbb{R}^3$ and the short rate $r(\mathbf{x}, t)$ is a sum of the first two components

$$r(\mathbf{x}, t) = x_1(t) + x_2(t),$$

which corresponds to setting $\rho_0 = 0$ and $\boldsymbol{\rho}_1^\top = [\,1\ 1\ 0\,]$ in the affine short rate formula (8). The risk-neutral state dynamics are governed by a SDE

$$\mathrm{d}\mathbf{x}(t) = -\boldsymbol{\mathcal{K}}^{\mathcal{Q}}\mathbf{x}(t)\,\mathrm{d}t + \boldsymbol{\Sigma}\,\mathrm{d}\mathbf{B}^{\mathcal{Q}},$$

where

$$\boldsymbol{\mathcal{K}}^{\mathcal{Q}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

with $\lambda \in \mathbb{R}$ and $\sigma_{ij} \in \mathbb{R}_+$ and $\boldsymbol{\mu}^{\mathcal{Q}}$ is set to zero. The diffusion coefficient $\boldsymbol{\Sigma}$ is time-invariant which means that the model volatility is constant (not state dependent). As a result, the essentially affine risk premium (17) is

$$\boldsymbol{\lambda}(\mathbf{x}) = \boldsymbol{\phi} + \boldsymbol{\Phi}\,\mathbf{x}(t),$$

leading to following $\mathcal{P}$-dynamics of the state

$$\mathrm{d}\mathbf{x}(t) = \boldsymbol{\mathcal{K}}^{\mathcal{P}}(\boldsymbol{\mu}^{\mathcal{P}} - \mathbf{x}(t))\,\mathrm{d}t + \boldsymbol{\Sigma}\,\mathrm{d}\mathbf{B}^{\mathcal{P}},$$

where $\mathcal{K}^{\mathcal{P}} \in \mathbb{R}^{3\times 3}$ and $\boldsymbol{\mu}^{\mathcal{P}} \in \mathbb{R}^3$. We refer to the *correlated* AFNS model parametrization, when $\boldsymbol{\Sigma}$ is triangular and $\mathcal{K}^{\mathcal{P}}$ is a full matrix. In the *independent* parametrization both of these matrices are diagonal.

For both state evolution and state estimation purposes, see Sections 3.2 and 4.2.1, we need to solve the equations (48), that are presented with ATSM dynamic model parametrization in (19). The following solutions are obtained for the independent AFNS model, where $\mathcal{K}^{\mathcal{P}}$ and $\boldsymbol{\Sigma}$ are diagonal

$$\begin{aligned}\mathbf{A} &= \exp[-\boldsymbol{\mathcal{K}}^{\mathcal{P}}\,(t_k - t_{k-1})]\\ \mathbf{A}_{ii} &= \exp[-\boldsymbol{\mathcal{K}}^{\mathcal{P}}_{ii}(t_k - t_{k-1})], \quad i = 1,2,3\end{aligned}$$

and

$$\begin{aligned}\mathbf{Q} &= \int_{t_{k-1}}^{t_k} \exp\left(-\boldsymbol{\mathcal{K}}^{\mathcal{P}}\,(t_k - s)\right)\,\mathbf{L}\,\mathbf{L}^{\top}\,\exp\left(-\boldsymbol{\mathcal{K}}^{\mathcal{P}}\,(t_k - s)\right)^{\top}\,\mathrm{d}s\\ \mathbf{Q}_{ii} &= \frac{\sigma_{ii}^2}{2\boldsymbol{\mathcal{K}}^{\mathcal{P}}_{ii}}\left(1 - \exp[-2\boldsymbol{\mathcal{K}}^{\mathcal{P}}_{ii}(t_k - t_{k-1})]\right), \quad i = 1,2,3.\end{aligned}$$

5.1.1 Measurement Model for Yields

The AFNS model solution for zero-coupon bond prices (10) is of the form

$$P(t,T) = e^{a(t,T,\mathbf{0}) + \mathbf{b}(t,T,\mathbf{0})^{\top}\mathbf{x}(t)}, \tag{62}$$

where $a(t,T,0)$ and $\mathbf{b}(t,T,\mathbf{0})$ are the time-t solutions to the ODE system (13) with boundary conditions defined by the 2nd and 3rd argument. With the AFNS model specifications, the ODE system (13) simplifies to

$$\begin{aligned}\frac{\mathrm{d}a(t)}{\mathrm{d}t} &= -\frac{1}{2}\sum_{i=1}^{3}\left(\boldsymbol{\Sigma}^{\top}\,\mathbf{b}(t)\,\mathbf{b}(t)^{\top}\,\boldsymbol{\Sigma}\right)_{ii}\\ \frac{\mathrm{d}\mathbf{b}(t)}{\mathrm{d}t} &= \boldsymbol{\rho}_1 + (\boldsymbol{\mathcal{K}}^{\mathcal{Q}})^{\top}\,\mathbf{b}(t),\end{aligned} \tag{63}$$

where $\boldsymbol{\rho}_1^\top = [\,1\;1\;0\,]$. Solving $a(t,T,0)$ and $\mathbf{b}(t,T,\mathbf{0})$ from (63) yields

$$\begin{aligned} a(t,T,\mathbf{0}) &= \frac{1}{2}\sum_{i=1}^{3}\int_t^T \left(\boldsymbol{\Sigma}^\top\,\mathbf{b}(s,T,\mathbf{0})\,\mathbf{b}(s,T,\mathbf{0})^\top\,\boldsymbol{\Sigma}\right)_{ii}\,\mathrm{d}s \\ \mathbf{b}_1(t,T,\mathbf{0}) &= -(T-t) \\ \mathbf{b}_2(t,T,\mathbf{0}) &= -\frac{1}{\lambda}\left(1-e^{-\lambda(T-t)}\right) \\ \mathbf{b}_3(t,T,\mathbf{0}) &= (T-t)e^{-\lambda(T-t)} - \frac{1}{\lambda}\left(1-e^{-\lambda(T-t)}\right), \end{aligned} \tag{64}$$

where analytical $a(t,T,\mathbf{0})$ is given in Appendix A.1. Measurement model for time-t yields $\mathbf{Y}(t,T) \in \mathbb{R}^d$, corresponding maturities $T_{1,2,\ldots,d}$, is then

$$\mathbf{Y}_i(t,T_i) = -\frac{a(t,T_i,\mathbf{0})}{T_i-t} - \frac{\mathbf{b}(t,T_i,\mathbf{0})^\top\mathbf{x}(t)}{T_i-t}. \tag{65}$$

5.1.2 Measurement Model for Caplet Prices

We derive here a closed form solution for caplet prices with the AFNS model. Let us consider the present time $t=0$ price of a caplet $V_{\mathrm{CPL}} \triangleq V_{\mathrm{CPL}}(K_c,T_{i-1},T_i)$, which has a strike $K_c \in \mathbb{R}_+$ and the reset and payment dates are T_{i-1} and T_i, respectively. A general transformation analysis based formula (35) for pricing such a caplet with affine term structure models was derived in the section 3.4.3; we restate it here for convenience

$$V_{\mathrm{CPL}} = G(\mathbf{0},\mathbf{b}_{T_{i-1}}) - (1+\tau K_c)e^{a_{T_{i-1}}}\,G(\mathbf{b}_{T_{i-1}},\mathbf{b}_{T_{i-1}}). \tag{66}$$

We denote with

$$G(\mathbf{u},\mathbf{q}) \triangleq G(\mathbf{u},\mathbf{q},\tilde{c},\mathbf{x}_0,T_{i-1})$$

the transformation $G(\cdot)$ that is defined in the equation (25). Additionally, we have $\tilde{c} = \ln\left(\frac{1}{1+\tau K_c}\right) - a_{T_{i-1}}$, τ is the year fraction $\tau = T_i - T_{i-1}$ and the pair of ODE system (63) solutions are denoted with $\mathbf{b}_{T_{i-1}} \triangleq \mathbf{b}(T_{i-1},T_i,\mathbf{0})$ and $a_{T_{i-1}} \triangleq a(T_{i-1},T_i,\mathbf{0})$. These solutions are given in (64).

To compute (66) in closed form, we need to solve the transformation

$$G(\mathbf{u},\mathbf{q}) = \frac{1}{2}\Gamma(\mathbf{u},\mathbf{x}_0,0,T_{i-1}) - \frac{1}{\pi}\int_0^\infty \frac{1}{v}\mathrm{Im}\left[\Gamma(\mathbf{u}+iv\mathbf{q},\mathbf{x}_0,0,T_{i-1})\,e^{-j\tilde{c}v}\right]\,\mathrm{d}v. \tag{67}$$

The first term in (67) is defined by the mapping (22), that is

$$\Gamma(\mathbf{u}, \mathbf{x}_0, 0, T_{i-1}) = e^{a(0,T_{i-1},\mathbf{u}) + \mathbf{b}^\top(0,T_{i-1},\mathbf{u})\,\mathbf{x}_0}, \tag{68}$$

where $a(0,T,\mathbf{u})$ and $\mathbf{b}(0,T,\mathbf{u})$ are the following solutions to the system (63), with boundary condition $\mathbf{b}(T) = \mathbf{u}$,

$$\begin{aligned}
a(t,T,\mathbf{u}) &= \frac{1}{2}\sum_{i=1}^{3}\int_t^T \left(\boldsymbol{\Sigma}^\top \mathbf{b}(s,T,\mathbf{u})\,\mathbf{b}(s,T,\mathbf{u})^\top\,\boldsymbol{\Sigma}\right)_{ii} \mathrm{d}s \\
\mathbf{b}_1(t,T,u_1) &= -(T-t) + u_1 \\
\mathbf{b}_2(t,T,u_2) &= -\frac{1}{\lambda}\left(1 - (1+u_2\lambda)e^{-\lambda(T-t)}\right) \\
\mathbf{b}_3(t,T,u_3) &= \left((T-t)(1+u_2\lambda) + u_3\right)e^{-\lambda(T-t)} - \frac{1}{\lambda}\left(1 - e^{-\lambda(T-t)}\right).
\end{aligned} \tag{69}$$

An analytical form for $a(0,T,\mathbf{u})$ is presented in Appendix A.2.

Now, consider the second term in (67). Assuming a complex boundary condition for the system (63), $\mathbf{b}(T) = \mathbf{z} \triangleq \mathbf{u} + jv\mathbf{q}$, where j is the imaginary unit, $v \in \mathbb{R}$ and $\mathbf{u}, \mathbf{q} \in \mathbb{R}^3$, the solutions (69) can be conveniently written as

$$a(0,T,\mathbf{z}) = -\alpha_3 v^2 + \alpha_2 v j + \alpha_1, \quad \mathbf{b}(0,T,\mathbf{z}) = \boldsymbol{\beta_2} v j + \boldsymbol{\beta}_1,$$

where $\alpha_i \in \mathbb{R}$, $\boldsymbol{\beta}_i \in \mathbb{R}^3$ and $\alpha_3 > 0$. We can make use of this by writing

$$\Gamma(\mathbf{u} + iv\mathbf{q}, \mathbf{x}_0, 0, T_{i-1}) = e^{-\alpha_3 v^2 + \alpha_2 v j + \alpha_1 + \boldsymbol{\beta}_2^\top \mathbf{x}_0 v j + \boldsymbol{\beta}_1^\top \mathbf{x}_0}.$$

Transforming this complex exponential into polar form, by applying Euler's formula, allows use to write the integrand in (62) as

$$\frac{1}{v}\mathrm{Im}\left[\Gamma(\mathbf{u} + jv\mathbf{q}, \mathbf{x}_0, 0, T_{i-1})\,e^{-j\tilde{c}v}\right] = \frac{1}{v}e^{-\alpha_3 v^2 + \alpha_1 + \boldsymbol{\beta}_1^\top \mathbf{x}_0}\sin(\alpha_2 + \boldsymbol{\beta}_2^\top \mathbf{x}_0 - \tilde{c}v),$$

which yields to a nice integral solution (only converges when $\alpha_3 > 0$)

$$\begin{aligned}
&\frac{1}{\pi}\int_0^\infty \frac{1}{v}e^{-\alpha_3 v^2 + \alpha_1 + \boldsymbol{\beta}_1^\top \mathbf{x}_0}\sin(\alpha_2 + \boldsymbol{\beta}_2^\top \mathbf{x}_0 - \tilde{c}v)\,\mathrm{d}v \\
&= \frac{1}{2}e^{\alpha_1 + \boldsymbol{\beta}_1^\top \mathbf{x}_0}\mathrm{Erf}\left(\frac{\alpha_2 + \boldsymbol{\beta}_2^\top \mathbf{x}_0 - \tilde{c}}{2\sqrt{\alpha_3}}\right).
\end{aligned}$$

To sum up, the measurement model for caplet prices in the AFNS model is

the formula (66). This formula can be evaluated in closed form.

5.2 Modeling Swap Exposure

We consider here exposure modeling of an interest rate swap with the AFNS model. The model is first calibrated to historical yield curve data; the historical performance and implied future behavior of the model are inspected. Then future yield curve scenarios are simulated in which the swap is priced; the value distributions and derived exposure measures are presented. Market information available up to the date 26.10.2012 (dd/mm/yyyy) is used in exposure modeling.

The swap considered is a standard 20-year payer-swap, where the fixed rate is set to 1.5 % and paid semi-annually. The floating rate, linked to the six month euribor, is received semi-annually. For simplicity, the floating rate is always fixed on the previous payment date to the prevailing six month zero-coupon yield, both legs have same payment dates and the year fraction between the dates is 0.5. The contract starts on 26.10.2012 (the first reset), matures in 20 years and has a 10 meur notional, which is not exchanged.

The AFNS model with independent parametrization is calibrated to historical zero-coupon bond yields of maturities $T = \{0.5, 1, 2, 3, 5, 7, 10, 15, 30\}$ given in years. That is, we define the measurement model (65) to encompass yields corresponding these maturities, and compute the MAP estimate (40) for model parameters with the help of linear continuous-discrete Kalman filter presented in the section 4.2.1. The measurement noise variance is fixed to 10^{-6} for each maturity and the observations are in the decimal form, i.e. $1\% = 0.01$. Uniform prior is used for all parameters; the prior for $\boldsymbol{\mu}^{\mathcal{P}}$ is defined across the whole real axis and for the rest of the parameters across the positive side of the real axis. The MAP optimization problem is solved using ready-made algorithms of the Mathworks MATLAB software.

The historical dataset used for parameter estimation consist of weekly yield observations from the period of $7.2.2003 - 26.10.2012$ totaling to 507 observations per maturity; because zero-coupon bonds are not actively traded as such, we use yields implied by the euro six-month swap curve. The dataset is obtained from the financial data provider ICAP.

5.2.1 Results

MAP estimates obtained for the independent AFNS model are shown in Table 1. The optimization routine converges to these estimates from various starting points, however, other optima, leading to similar state estimates with similar objective function values in the optimum, are also found.

Table 1: MAP estimates for the independent AFNS model parameters obtained from historical yield curve estimation. The MAP estimation objective function (39) value in the optimum is $-27 \cdot 10^3$.

i	$\boldsymbol{\mathcal{K}}^{\mathcal{P}}_{ii}$	$\boldsymbol{\mu}^{\mathcal{P}}_{i}$	$\boldsymbol{\Sigma}_{ii}$	λ
1	0.1521	0.0489	0.0051	0.4447
2	0.2212	-0.0285	0.0067	
3	1.0000	-0.0275	0.0165	

Running the linear Kalman filter through the historical yield observations with the AFNS model specifications and above parameter values allows us to compare the model implied yield curves to the real observed ones. In Figure 6 are shown the AFNS model implied and the real observed curve on the last observation date 26.10.2012 alongside the deviation (error) of the model fit $\hat{y}_i$ from the observations y_i, that is $(y_i - \hat{y}_i)$, for maturities i that are direct model outputs. The magnitude of error is measured in basis points; 1 bp is a 100^{th} of a percentage point.

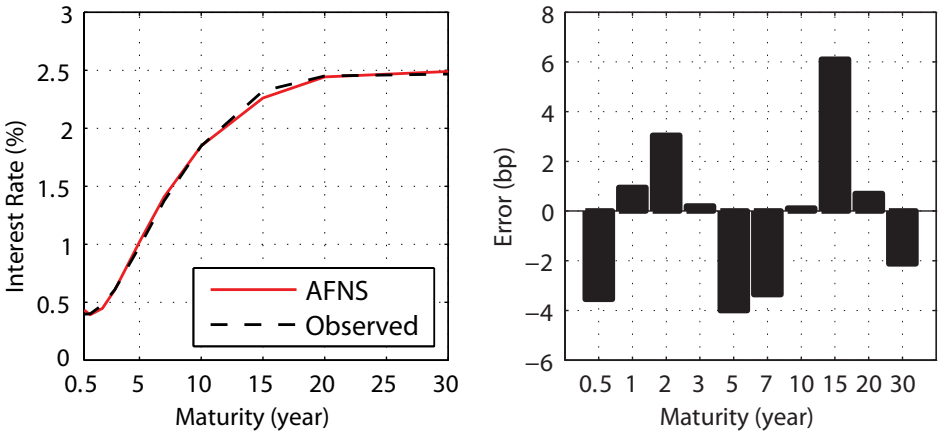


Figure 6: On the left, the AFNS model implied and the real observed yield curve on the last calibration date. On the right, errors of the model fit measured in basis points (bp); 1bp is a 100^{th} of a percentage point.

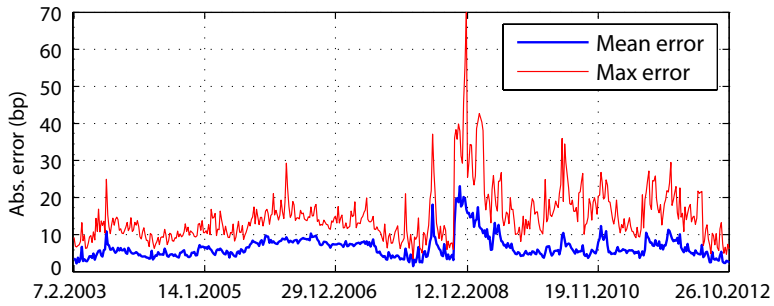


Figure 7: Mean and maximum absolute errors of the AFNS model yield curve fit throughout the calibration dataset.

The deviation of the AFNS model yield curve from the real observed one on the last calibration date 26.10.2012 is small across the different maturities, as can be seen in Figure 6. However, the model fit is not that precise throughout the historical estimates; 7 shows the mean and maximum of absolute model error, $|y_i - \hat{y}_i|$, computed across the maturity dimension on each calibration date. The mean error is fairly consistently less than 10 bp excluding the Fall of 2008 and Spring of 2009, and the maximum error is mostly less than 20 bp, although it is more volatile after year 2008.

A breakdown of the model error in the maturity dimension is presented in Table 2; error statistics for each maturity are computed across the time dimension and shown in absolute $|y_i - \hat{y}_i|$ and relative terms $\frac{|y_i - \hat{y}_i|}{y_i}$. Both mean error measures are small and stable across the maturities; quantiles for relative error imply slightly longer tails for error distributions of the 30-year maturity and the maturities below five years than for the maturities from 5 years to 20 years. Still, there are no great discrepancies in estimation precision of different maturities.

Table 2: Absolute error statistics of historical yield estimates across maturities. The mean and a 95 % quantile ($Q_{95\%}$) of absolute errors are given both in absolute and relative terms, i.e. in basis and percentage points.

	Maturity (year)									
Abs. error	0.5	1	2	3	5	7	10	15	20	30
Mean (bp)	6	5	8	6	3	5	7	8	4	13
$Q_{95\%}$ (bp)	15	11	18	12	7	11	14	19	13	26
Mean (%)	3	2	3	3	1	2	2	2	1	3
$Q_{95\%}$ (%)	7	8	8	7	3	3	4	5	3	9

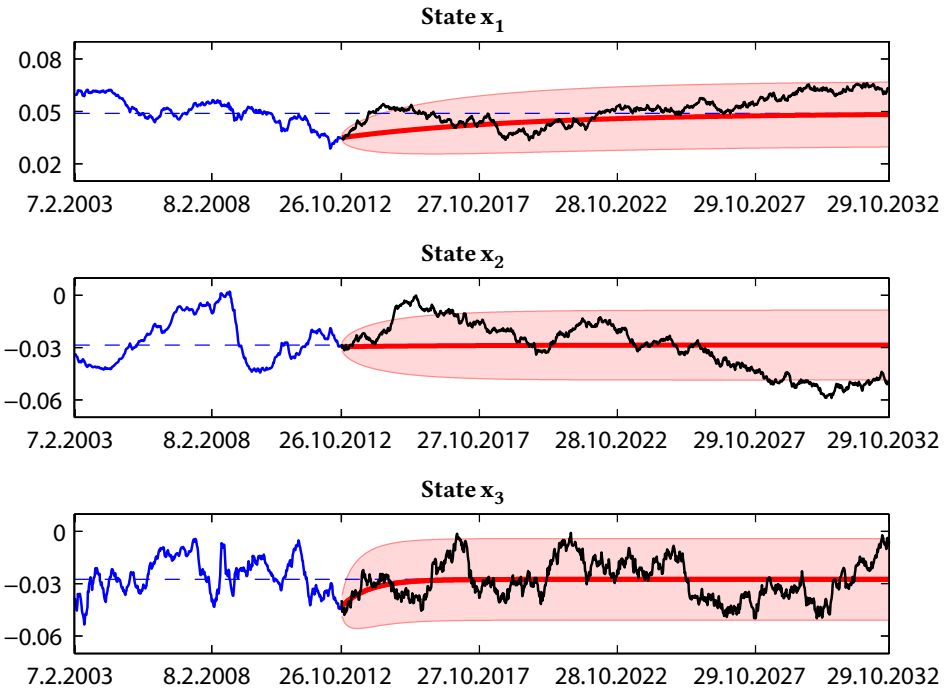


Figure 8: The filtered historical state evolution and the model implied future distributions for state variables. The filtered state mean is shown in solid blue and the long-run mean parameter $\boldsymbol{\mu}^{\mathcal{P}}$ in dashed blue. The model implied future state mean is depicted in dark red with a 90 % confidence interval in light red. A simulated sample of future state evolution is shown in black.

The AFNS model implied future state behavior (under the $\mathcal{P}$ measure) is illustrated in Figure 8 by contrasting the future state mean and variance with the historical evolution of state mean filtered from the calibration dataset observations. The future state mean (in dark red) reverts smoothly back to the long-run mean implied by the parameter $\boldsymbol{\mu}^{\mathcal{P}}$ (in dashed blue), which corresponds roughly to the average of historical state mean estimates (in solid blue). Variance around the mean is presented with the help of 90 % confidence interval (in light red); the width of the confidence intervals grows first and then saturates to a stable value, which is within the boundaries of historical state estimate movements. Although the state mean evolution is smooth, a sample state path (in solid black) resembles the jagged evolution of the filtered mean.

To illustrate the model implied future yields, Figure 9 shows yield distributions in one and ten years of time from the last observation date 26.10.2012. Yields are computed for maturities from 0.5 year to 30, year with 0.1 year

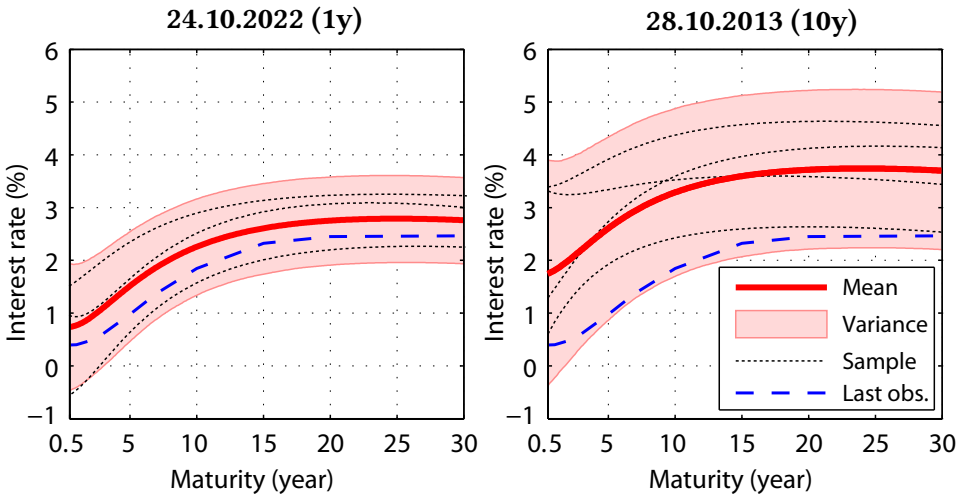


Figure 9: Mean and variance (90 % confidence interval) of future yields with a few sample curves and the last observed curve (on 26.10.2012).

spacing from 10000 state simulations and the summary statistics are obtained for each maturity separately from the simulations.

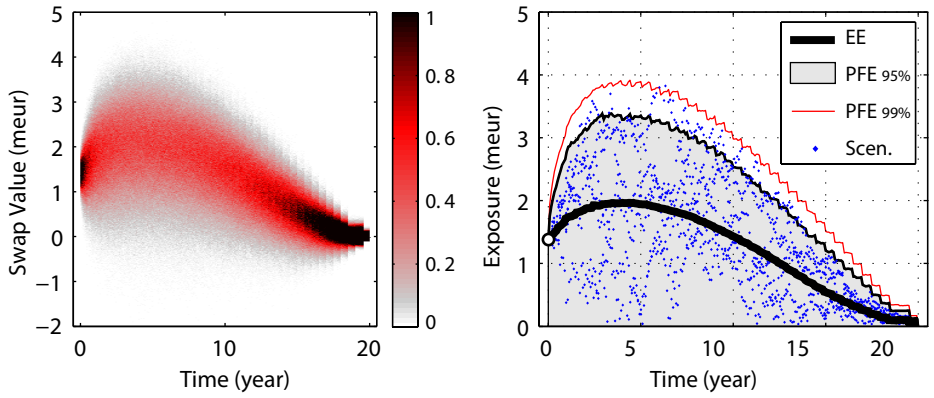
Several observation can be made from the future yield curve distributions in Figure 9. There is clearly more variance in the simulated yields after ten years than after a year. This corresponds to the fact that the variance of state variables is greater after ten years than after a year, as can be seen in Figure 8. The overall level of the yields is also higher for the ten-year yields than for the one-year yields — the interpretation of the state variable x_1 as the level of the yield curve indicates that the model implied interest rates rise from the initial level, on average, for the first 15 simulation years. This is evident from the mean yields, the shape of which is very similar between the simulation times. Finally, the AFNS model allows for negative yields; this is a property of all Gaussian type of term structure models. We have treated negative yields in this case study by capping the maximum zero-coupon bond price (discount factor) to one.

After model calibration and inspection, we apply it to exposure estimation of the 20-year payer-swap. We use monthly spaced exposure dates from the start date of the contract 26.10.2012 to its maturity in 20 years, which amounts to total of 240 exposure dates. For each exposure date, excluding the start and end dates, we simulate 10000 state scenarios under the

physical measure with the AFNS model. Given the state values and model parameters, the swap can be valued in each scenario.

Figure 10a shows a map of the swap future values. The value map is obtained by pricing the trade in simulated state scenarios and mapping these values to a grid, that has monthly spacing in time dimension and 10 keur in the value dimension. Cells between the grid points are colored according to the percentage of scenarios having values between the corresponding grid point values; the scale has been capped to one, i.e. cell values above one are set to one. The swap is valued at time $t = 0$ (26.10.2012) with the AFNS model yield curve to 1.4 meur, which is roughly 1 % lower price than the one computed from the observed curve. The initial value is positive, because the swap fixed rate, which is paid in payer-swap, is lower than the swap rate is. Due to the AFNS model calibration implying a general tendency of rising interest rate simulations, and the swap being a payer-swap (floating rate received), the swap value map is skewed towards positive trade values.

Time-dependent exposure measures derived from the value scenarios are shown in Figure 10b. All the exposure measures start from the swap initial value (which is known), rise for a while and then start to decrease reaching zero at the contract maturity. The potential future exposure (PFE) measures are clearly higher than the expected exposure (EE). The difference between 95 % and 99 % confidence levels is not very big, though.



(a) Swap value map **(b)** Exposure measures

Figure 10: In panel (a) is illustrated the distributions of swap value scenarios on each exposure date. In panel (b) is shown exposure measures derived from the value scenarios and some demonstrative exposure scenarios.

5.2.2 Discussion

After the 2007–2008 financial crisis forward rates of swaps with a six-month tenor are computed from the corresponding tenor swap curve, but the discount factors are obtained from a different curve (EONIA). In the single curve pricing framework based treatment we have used interest rates of the eur six-month swap curve as proxies for zero-coupon bond yields, calibrated the model to historical observations of the swap curve and used the model for exposure modeling of a swap that has a six-month tenor. Discounting the swap value with six-month based yields underestimates the exposure, because it undervalues the swap. Additionally, to model the exposure of a swap with different tenor we would need to simulate the evolution of yields implied by the corresponding swap curve, so that the used forward curve would be the correct one. Hence, it would be of great interest to study the historical performance of a model calibrated to six-month yields with, e.g., three month yield data to see, if same parameter estimates could be used.

The choice to calibrate the AFNS model with yields of maturities starting from 0.5-year pre-empts calibration problems related to the very short end of the yield curve, which usually exhibits different dynamics and more volatile behavior compared to the rest of the curve. Because the used exposure dates have monthly spacing, these yields are still needed in the exposure modeling. We have not inspected the model implied yields for maturities less than four months, because the impact of these yields is generally negligible due to their short time to maturity.

The variety of historical yield curves used in the model calibration is significant, because the calibration dataset includes roughly five years of pre- and post-financial crisis time, and the months of turbulence during 2007–2008. The AFNS model is able to capture the observed curves successfully. This is evident from the small overall historical errors, see Figure 7, and average absolute errors less than 10 bp per maturity, see Table 2. The relative mean errors further indicate that the yields of different maturities are estimated with same relative precision, on average. For exposure modeling a rough historical estimation precision of a model is sufficient, because the interest is in future yield curves. Although, the worse the fit to the prevailing curve the bigger the difference between initial exposure and the real swap value.

The future state evolution under the physical measures implied by the AFNS model resembles the estimated historical evolution, as can be seen from the mean and confidence intervals in Figure 8 — without the graphical cues, the sample state path would be hard to distinguish from the estimated state mean. In this light, the model implied future yield curves, seem both credible and, from the exposure modeling point of view, encouraging, because the model is able to produce in the future the variety of yield curve shapes present in the calibration dataset. However, the Kalman filter allows the estimated state variables to be correlated regardless of the independent model parametrization. The correlation is not strong in this case study however, and hence we are comfortable with using the independent AFNS model instead of the correlated one.

We regard the independent AFNS model to be a valid candidate for swap exposure modeling on the grounds of historical estimation performance and implied future yield curve evolution, although more detailed study of both aspects should be conducted in real exposure modeling endeavor. The exposure profile for the 20-year payer-swap obtained with the model is shown in Figure 10b; the profile corresponds both to intuition and textbook examples of a swap exposure profile. In general terms, the exposure increases initially due to flexible yield curve simulations and because, by the exposure definition, the negative value scenarios are treated as zeros. After the first five years the exposure measures start to decrease as the number of remaining cash flows decreases and this effect becomes more pronounced; from the swap value map in Figure 10a can be seen how the spreads of value scenario distributions narrow down towards the end of the contract. The jaggedness of the PFE curves reflect the payment-reset date cycle of the swap.

Further research should be carried out to assess validity of the obtained exposure profiles, their error bounds and sensitivity to model parameter estimates. In an ideal case the calibration dataset could be divided to training and test dataset, so that the exposure profiles could be benchmarked against the exposure of realized yield curve evolution. Due to long maturities of swaps this would require a huge dataset though. However, the exposure profiles could also be compared against exposures implied by different models; the model of (Rebonato *et al.*, 2005) is one particularly interesting candidate considering swaps.

5.3 Modeling Joint Exposure of a Cap and a Swap

In this case study, we seek to model the exposure of a derivative portfolio which consists of an interest rate cap and swap with the AFNS model. The overall progress of this study corresponds to that of the previous one: historical performance and implied future behavior of the model are inspected first and exposure computation steps are taken afterwards. The trades considered are forward starting and defined so that the cap hedges the interest rate risk of the swap; these specifications render this case study real-life trading situation oriented. The exposure is modeled based on market information available up to the date 26.10.2012.

The cap and swap are defined as follows. Both trades have the first reset date on 26.4.2013, mature in ten years and have a notional of 10 meur. The payment and reset dates are six months apart and the year fraction between these dates is a constant 0.5. Payments of both contracts are fixed to the reset date prevailing six month zero-coupon yield. With these specifications the cap is a 10-year six-month cap consisting of caplets with six-month tenor; the cap strike rate is set to 3 %. The swap is a 10-year receiver swap (fixed rate is received, floating rate paid); we set the fixed rate to 1.88 %, which is the model implied fair swap rate for the contract on 26.10.2012.

The independent AFNS model is used for exposure estimation and it is calibrated both to zero-coupon bond yields and ATM cap prices; the cap prices are for maturities $T = \{3, 5, 7, 10\}$ with six month tenor and the yields are same as previously, that is, the yields for maturities $T = \{0.5, 1, 2, 3, 5, 7, 10, 15, 30\}$ (in years) are considered. Because the cap pricing formula is nonlinear in the state, we apply the continuous-discrete cubature Kalman filter presented in the section 4.2.2 for the MAP problem, which is solved again with the MATLAB software. The measurement noise variance is set to 10^{-6} for all observations and uniform prior is used for the parameters; for $\boldsymbol{\mu}^{\mathcal{P}}$ it is defined across the whole real axis and for the rest of the parameters across the positive side.

The dataset used in model calibration consist of weekly ATM cap prices, corresponding strike rates, and yields from the period of 7.2.2003 – 26.10.2012. The cap prices are obtained by transforming the six-month ATM cap im-

plied volatilities, obtained from data provider ICAP, to corresponding prices using the the Black-76 model (Black, 1976) and the known eur six-month swap curves, from which the ATM strike rates are also computed.

5.3.1 Results

We use the parameter estimates obtained in the previous case study, see Table 1, as an initial guess for the MAP optimization. The MAP estimation problem here is nonlinear and notably more burdensome computationally than in the linear case. Obtained MAP parameter estimates for the AFNS model, when both yields and cap prices are estimated, are shown in Table 3.

By comparing the estimates in Table 3 to the initial values in Table 1, it can be seen that the optimum is very similar to the initialization. Indeed, the energy function (39), which is minimized in MAP estimation, returns $-28.9 \cdot 10^3$ with the initial parameter values, and $-30.1 \cdot 10^3$ with the presented optimum — a random initialization returns $+200 \cdot 10^3$, for example — indicating that the initial parameter values are already near a local optimum; this is further confirmed by rapid convergence of the optimization.

Table 3: MAP estimates for the independent AFNS model parameters obtained from joint estimation of historical yields and ATM cap prices; the energy function returns $-30.1 \cdot 10^3$ with these parameter values.

i	$\boldsymbol{\mathcal{K}}_{ii}^{\mathcal{P}}$	$\boldsymbol{\mu}_i^{\mathcal{P}}$	$\boldsymbol{\Sigma}_{ii}$	λ
1	0.1512	0.0533	0.0051	0.5563
2	0.1643	-0.0260	0.0059	
3	0.1849	-0.0302	0.0128	

We study the model historical performance by comparing the model implied yield curves and cap prices to the observations used in the calibration. For this, we need the historical states estimates that are obtained by running the continuous-discrete cubature Kalman filter through the calibration dataset with the AFNS model specifications and above parameter values. The yield estimates are also compared to the estimates of the previous case study.

The model implied yields and cap prices on the last observation date (26.10.2012) are compared to the observed ones in Figures 11 and 12. Regarding the yield curve estimate, the short-end, maturities below 5-year,

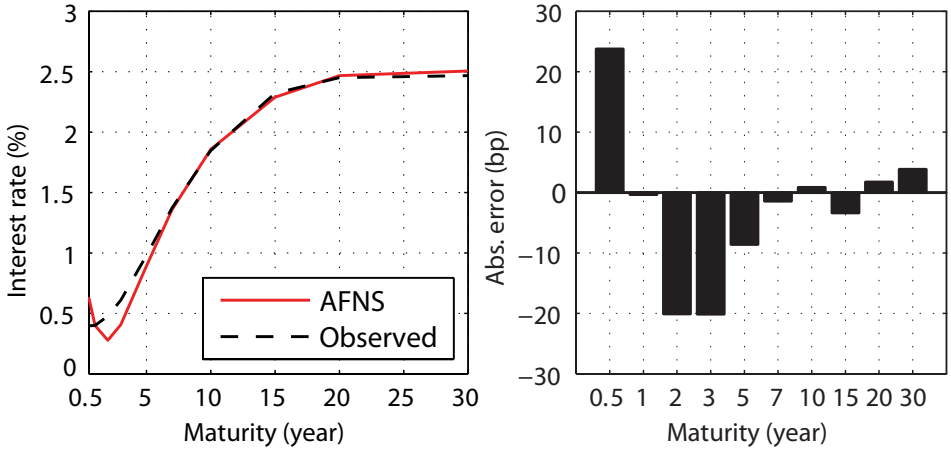


Figure 11: The AFNS model implied and the real observed yield curve alongside the model error for calibration maturities on 26.10.2012.

is badly off both in qualitative and quantitative terms; the errors are large is absolute terms, huge in relative terms and instead of a monotonically decreasing yield curve the estimate features a U-shape in the short-end.

Estimates for ATM cap prices (notional 1 eur, strikes given in Table 4) on the last calibration date 26.10.2012 are qualitatively good, shape of the price curve is captured, and all the relative errors are below 10 %-points. The model error is given in relative terms $\frac{(\hat{y}_i - y_i)}{y_i}$ for caps, because it turns directly into mispricing, i.e. an error of +10 % means overvaluation by 10 %.

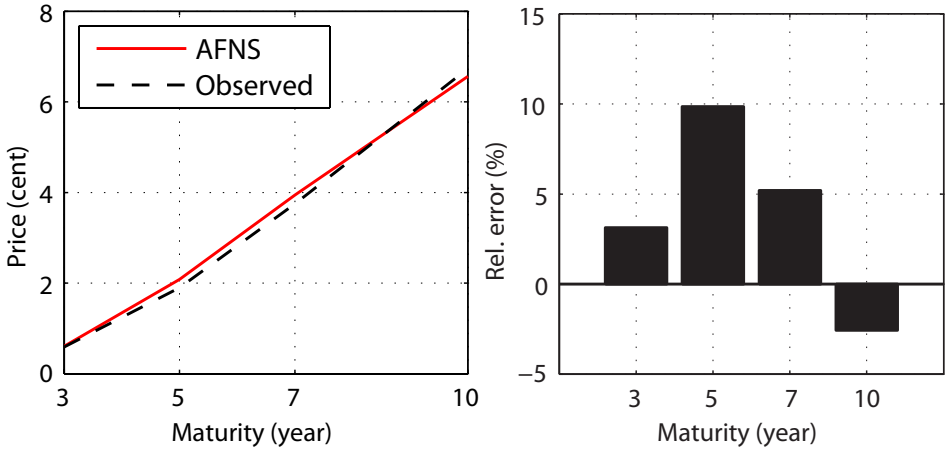


Figure 12: The AFNS model implied and the real observed ATM cap prices with relative errors of the model fit for calibration maturities on 26.10.2012.

Table 4: Computed ATM strikes for caps on 26.10.2012.

Maturity (year)	3	5	7	10
Strike (%)	0.65	1.04	1.43	1.88

Naturally, cap prices with other than ATM strike rates can also be obtained with the model; a computed cap price surface for 26.10.2012, in maturity and strike dimensions, is shown in Figure 13. The surface is formed of cap prices that are computed on a grid with spacing $1:1:10$ (year) in maturity and $0.5:0.5:5$ (%) in strike dimension. Spacing $a:b:c$ denotes the points from a to c with constant interval b. The surface is sound in the sense that the cap prices decrease when strike increases or maturity shortens.

Measures for the AFNS model historical estimation error throughout the calibration dataset are presented maturity-wise in Tables 5 and 6 separately for yields and cap prices; the mean and 95 % quantile ($Q_{95\%}$) of absolute errors are presented both in absolute are relative terms. Time-evolution of the overall absolute error statistics, computed across the maturities on each calibration date, are shown for yields and cap prices in Figure 14.

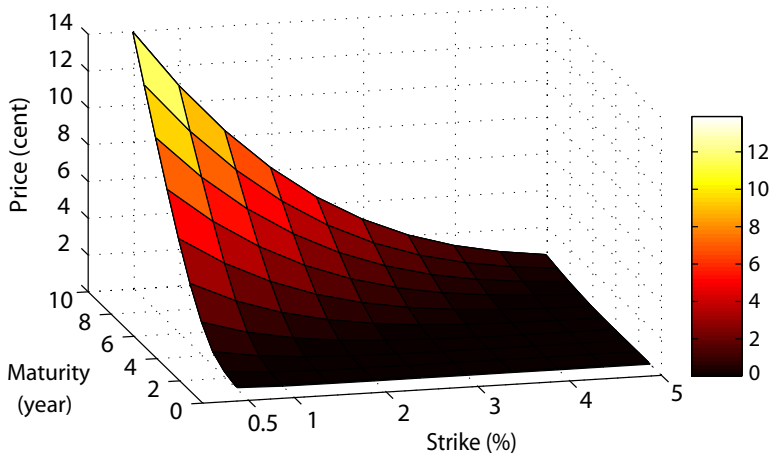


Figure 13: A mode implied cap price surface on 26.10.2012.

The yield estimate error statistics in Table 5 are relatively stable for maturities from one year to 20 years, although they are roughly double in magnitude compared to corresponding statistics of previous case study, see Table 2. The yield estimates for maturities 0.5 year and 30 year however, differ from the rest with higher average errors and error quantiles that indicate

longer tails for error distributions. The higher average error is also evident from the overall error measure evolution shown in Figure 14a. The maximum error explodes in late 2008 and often peaks very high in 2009–2012. The former is due to the emergence of very atypical yield curves shapes and the latter corresponds to situations where the 30-year yield is notably lower than the 20-year yield; the AFNS model with cap prices is not flexible enough to capture these yield curve shapes.

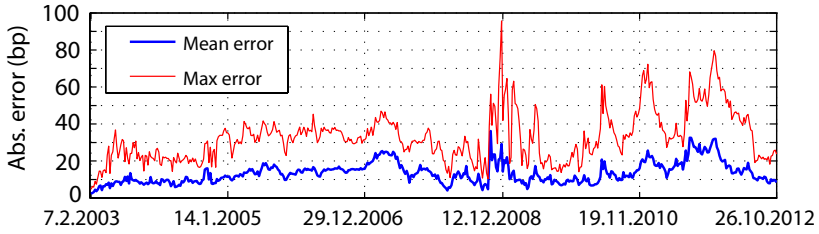
The relative error statistics for ATM caps with maturities 5, 7 and 10 years are of the same magnitude as the yield error statistics indicating that, even though there are more yields to be estimated, these cap prices are estimated roughly with same relative precision, see Tables 5 and 6. This is not the case with the 3-year cap prices, for which the error statistics are clearly larger. The peaks of the overall cap price error measures in Figure 14b are due to radical and fast changes in the cap prices that are not captured right away by the model.

Table 5: Absolute error statistics of historical *yield estimates* for the calibration maturities. Mean error and 95 % quantile ($Q_{95\%}$) are presented both in basis and percentage points.

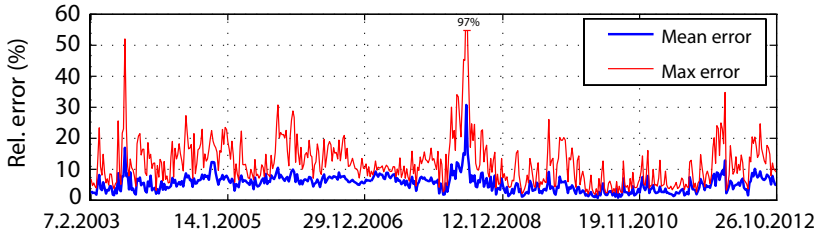
	Maturity (year)									
Abs. error	0.5	1	2	3	5	7	10	15	20	30
Mean (bp)	14	6	11	11	11	11	12	12	17	31
$Q_{95\%}$ (bp)	31	17	23	23	23	26	30	28	36	60
Mean (%)	7	4	6	5	4	3	3	3	4	8
$Q_{95\%}$ (%)	16	12	13	12	9	9	10	9	12	20

Table 6: Absolute error statistics of historical *ATM cap price estimates* for the calibration maturities. Mean error and 95 % quantile ($Q_{95\%}$) are presented both in absolute and relative terms.

	Maturity (year)			
Abs. error	3	5	7	10
Mean (cent $\cdot 10^{-2}$)	9	7	12	19
$Q_{95\%}$ (cent $\cdot 10^{-2}$)	21	21	25	40
Mean (%)	11	3	4	3
$Q_{95\%}$ (%)	23	9	9	7



(a) Yield estimate errors



(b) Cap price estimate errors

Figure 14: Mean and maximum estimation error of yields in panel (a) and cap prices in panel (b) computed across maturities on each calibration date.

Future scenarios for (as well as estimates of) yields and cap prices are based both on model specifications and simulated (estimated) states. The historical evolution of the estimated state mean is shown in Figure 15 alongside the model implied future state distributions, that are visualized with the help of mean and 90 % confidence interval. The long-run mean parameter $\boldsymbol{\mu}^{\mathcal{P}}$ and a sample of state evolution are also presented.

The state variables' future means revert to $\boldsymbol{\mu}^{\mathcal{P}}$ which is close to historical averages and the estimated historical state evolution is within the confidence intervals, see Figure 15. Both the historical and implied future state evolution are very similar to those of the previous case when compared to the Figure 8. We only notice differences with bare eyes in the state x_3 estimated historical evolution and confidence intervals.

We illustrate future yields and cap prices derived from the future state distributions and model specifications in Figures 16 and 17, respectively. The yields are computed in one and ten years of time from the last calibration date for maturities $0.5 : 0.1 : 30$ (year) from 10000 state simulations. Future price distributions of a cap with a constant 10-year maturity and 3 % strike are computed from 10000 state paths on a weekly interval for ten years and compared to historical estimates of corresponding cap prices.

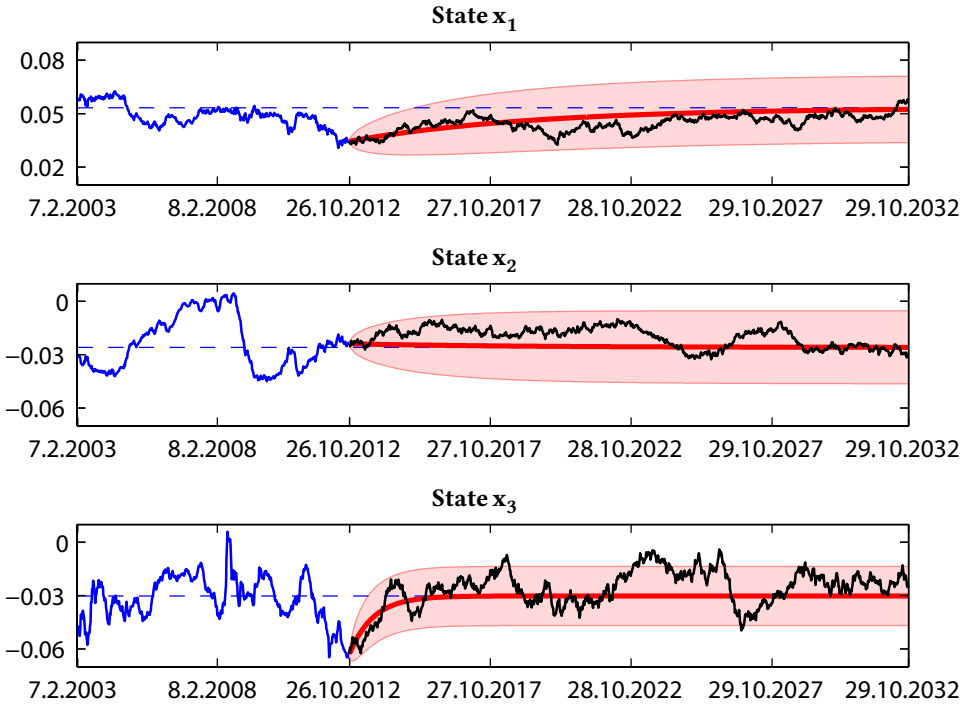


Figure 15: The filtered historical state evolution and the model implied future distributions for state variables. The filtered state mean is shown in solid blue and the long-run mean parameter $\boldsymbol{\mu}^{\mathcal{P}}$ in dashed blue. The model implied future state mean is depicted in dark red with a 90 % confidence interval in light red. A simulated sample of future state evolution is shown in black.

Figure 16 shows that the yields after ten years of simulation time, in the panel on the right, are higher in average and have higher variance than the yields after a year of simulation time in the left panel. The shape of mean yields is very similar between the simulation times and also with the last observed curve. The variance is somewhat greater for the very short maturities but, in general, of the same magnitude across the maturities. As is the case with the state evolution, these future yield distributions are very similar with the distributions obtained in previous case study, see Figure 9.

The cap price distribution evolution shown in Figure 17 has a rising mean with increasing variance. The variance is illustrated with the help of 90 % and 95 % confidence intervals; the difference between the two confidence intervals is quite large on the upside indicating, that the price distributions have fairly long tails to that direction. Compared to the historical cap prices, the future distributions are in the same ballpark.

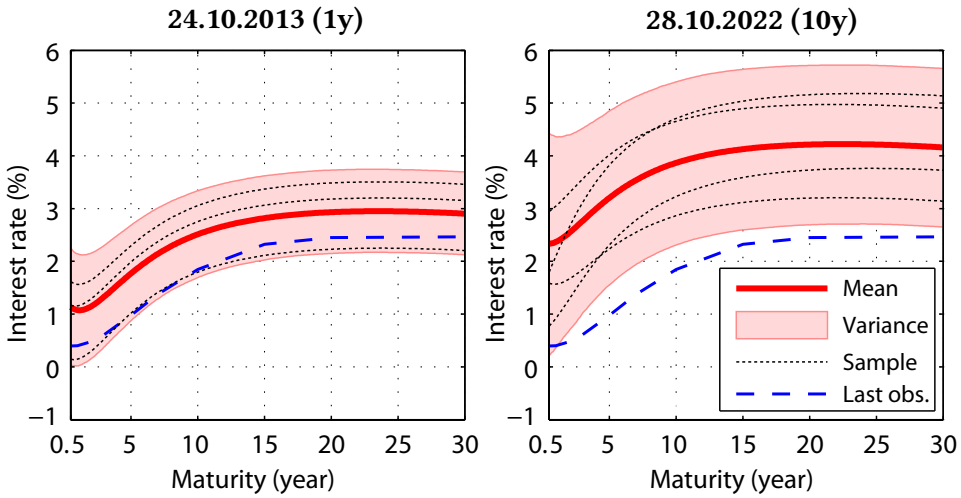


Figure 16: Mean and variance (90 % confidence interval) of future yields with a few sample curves and the last observed curve (on 26.10.2012).

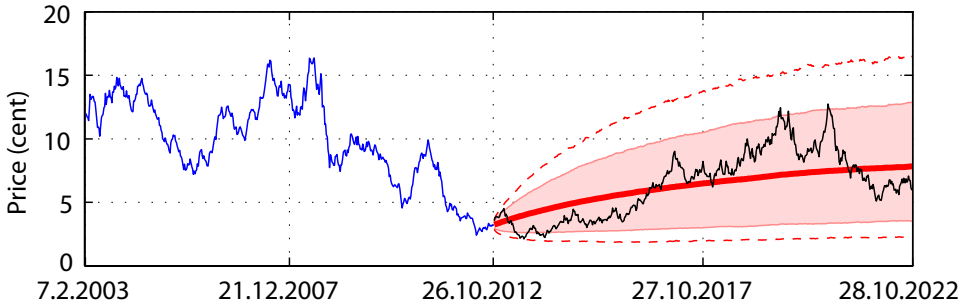


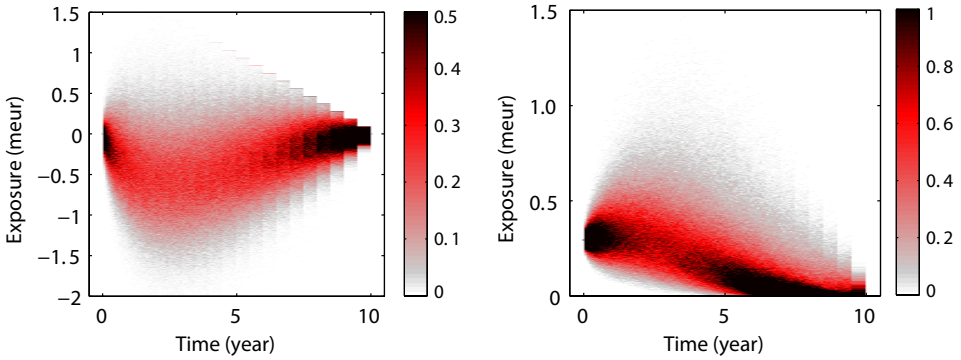
Figure 17: Estimated historical 10-year 3 % strike cap price (in blue) and future price distributions' mean (in dark red), 90 % and 95 % confidence intervals (in light and dashed red, respectively), and a sample price path (in black).

To compute exposure measures for the considered interest rate cap and swap, we simulate 10000 state paths under the physical measure with a monthly interval, that is monthly spaced exposure dates, up to the maturity of both contracts, which is in 10 years from the first exposure date 26.10.2012 and in nine and half years from the first reset date of both contracts 26.4.2013. In each simulated scenario both trades are valued given the state value, AFNS model and trade properties.

Maps of the obtained trade value distributions are shown in Figure 18 and the derived exposure measures in Figures 19 and 20. The value maps are formed by computing the number of trade value scenarios, on each exposure

date, in cells with a value dimension discrezation of $0 : 3\,\text{keur} : 1.5\,\text{meur}$ in the case of the cap and $-2\,\text{meur} : 3\,\text{keur} : 1.5\,\text{meur}$ with the swap. Both obtained maps are further normalized, for a better visualization, so that the cap grid cells containing more than 1 % of the total number of scenarios are set to value one and, in similar manner, the swap grid cells containing more than 0.5 % of the total number of scenarios are set to value 0.5.

The future swap values are clearly more scattered than the cap values, as can be seen in Figure 18, though the contracts have same notional amount and same number of payment dates. A swap can have a negative value where as a cap is an option with a minimum value of zero. The swap in question is a receiver-swap, and the obvious value distribution skew towards negative values is caused by the model-implied tendency for raising interest rates; the floating rate is paid, hence raising interest rates implies decreasing trade value. The cap is initially out of the money (strike 3 % compared to ATM strike of 1.88 %), but the trade value is still positive, as the definition for ATM cap is that its price equals the price of a corresponding floor. The cap value distributions spread out for four first years, but after two years the values on which the scenarios are concentrated start to decrease notably. Apparently, the interest rates rise, on average, slower than the cap goes even more out of the money due to shortening maturity and hence decreasing number of cash flows.

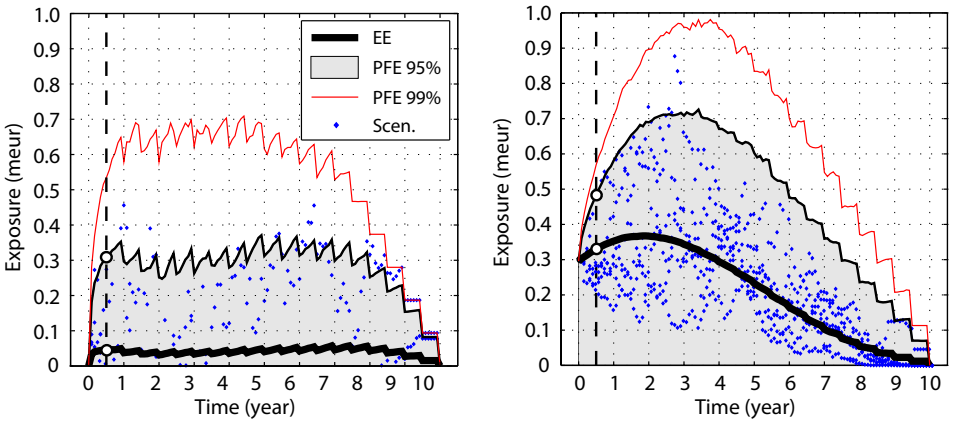


(a) Swap value map **(b)** Cap value map

Figure 18: Value scenario maps for the trades. Coloring indicates the percentage of scenarios on the corresponding cell. The color scales for both maps are capped, i.e., values exceeding the maximum scale value have been capped to the maximum value.

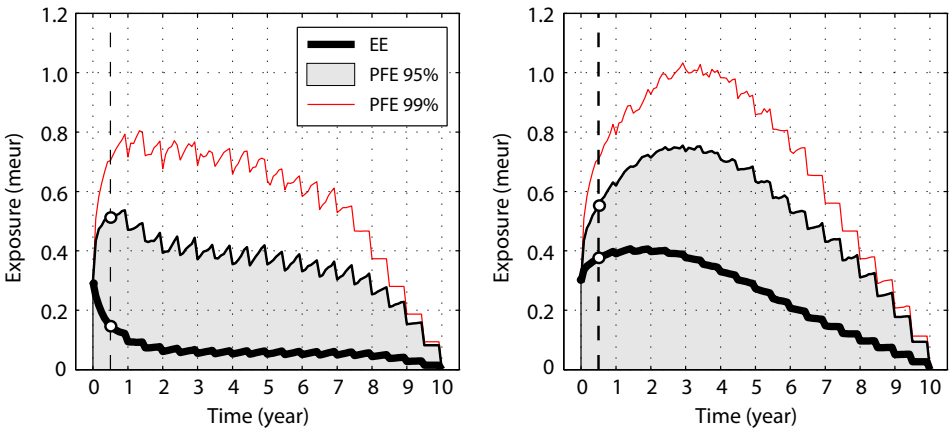
The obtained time-dependent exposure measures for the trades are presented in Figure 19. Though the swap values vary more, the estimated cap exposure is greater, measured with any of the exposure measures, because the swap values distributions are skewed towards negative values. For this reason, the expected exposure of the swap is relatively small. The number of sample value scenarios depicted in the figures is same for both trades, but in the swap case there are clearly fewer positive scenarios. The $\text{PFE}^{95\,\%}$ and $\text{PFE}^{99\,\%}$ curves for the swap see-saw notably, and the EE curve slightly, in the reset date payment date cycle. In the cap case the payment reset date cycle is evident only towards the end of the contract. The white dots at 0.5 year of time denote the EE and $\text{PFE}^{95\,\%}$ implied contract prices at their start date; both trades are forward starting, hence their value, and exposure, varies according to market variables after the contracts are struck.

Measures of portfolio level exposure are straightforward to derive from the value scenarios of the trades. Figure 20 shows the exposure measure for the portfolio, that consists of the considered cap and swap, with a netting agreement and without it no-netting. The netting agreement significantly reduces exposure; this is because the negative swap values set-off the exposure caused by the positive cap values. Exposure measures for the portfolio in the no-netting case are fairly similar to those of the cap, because the swap does not add much exposure to it.



(a) Swap exposure **(b)** Cap exposure

Figure 19: Time-dependent exposure measures for the 10-year cap and swap with a score of sample value scenarios.



(a) Netting **(b)** No-netting

Figure 20: Portfolio level exposure measures for the cap and swap with a netting agreement and without it.

5.3.2 Discussion

In this case study the AFNS model is calibrated to historical six-month ATM cap prices and to same yields as in the previous case. Motivation for using cap prices in the calibration stems from the possible advantages in model identification and state estimation (see, Christoffersen *et al.*, 2012). Also, when cap prices are needed as outputs of the model, it makes sense from the viewpoint of state space models to use them in training the model. Although, according to the results, the advantages of using cap prices in the model calibration are not so obvious; the objective function value in the MAP problem is of the same magnitude with the parameters that estimated solely using yields and those estimated with both cap prices and yields. However, one cannot draw even tentative conclusions based on this simplistic case study. There is not much previous research available — according to our knowledge and taking into account the context of evolutionary pricing of interest rate derivatives with ATSMs — possibly because the parameter estimation problem becomes nonlinear with the introduction of derivative prices as outputs. With the simplistic AFNS model an analytical solution for caplet prices can be derived facilitating the nonlinear parameter estimation. With more complex ATSMs this might not be so. All in all, we regard this topic to be of interest in future research.

The cap prices used in the model calibration are solved from a dataset of ATM implied volatilities and known eur six-month swap curves according to the single curve pricing framework principles (no EONIA discounting). As a result, the ATM cap prices used in the model calibration do not correspond exactly to those observed in the market. However, because the AFNS model cannot accommodate EONIA discounting, it is better to calibrate by using cap prices that are consistent with the model mechanics rather than with the EONIA discounting. Additionally, we disregard the prices corresponding to maturities below three years, because these prices correspond to three-month caplets in the euro-area.

From computational point of view, it would be more efficient to use caplet prices in the model calibration than cap prices. Additionally, using caplet prices would provide the components forming caps as targets rather than a summary value of them. The downside in using caplet prices is that the number of model outputs increases dramatically. The pros and cons of caplet vs. cap price calibration is a topic for further research. In this case study we have used ATM cap prices in the calibration. ATM prices are used rather than prices of fixed strikes, because they are not so sensitive to changes in the overall interest rate level; for example, a cap with 6 % strike is way out of the money on 26.10.2012 but it would have been in the money several years earlier. Still, in exposure modeling of a cap or floor the strike is naturally fixed, hence the model performance with varying strikes is also of interest.

Finally, we have calibrated the AFNS model to study joint exposure modeling of these derivatives with a tenor of six months. In this case, only the yields of the six-month swap curve are of interest in single curve based pricing. However, if the cap would have had a tenor of, say three months, then the cap prices are based yields and volatility of the three-month swap curve, where as the swap prices are based on the six-month curve yields. This is a problem with the AFNS model, and ATSMs in general, because they do not make it possible to model different yield curves jointly. It would be of great practical interest to study how to accommodate the need for modeling different yield curves simultaneously in portfolio exposure modeling. Disregarding EONIA discounting is acceptable, but using the wrong forward curve is not.

There seems to be a trade-off between the AFNS model historical estimation precision of yields and inclusion of cap prices in the measurement model. The estimation error statistics in Table 5 for yields, when cap prices are also estimated, indicate larger errors on average, especially for the 0.5-year and 30-year yields, and longer tails for the errors distributions (quantiles), than in the previous case, see Table 2, where the measurement model consists solely of yields. The cap prices are estimated roughly with the same relative precision as yields with maturities from 5 to 15 years, excluding the cap with shortest maturity, 3 years, that has worst relative error statistics of all model outputs, see Table 6. All in all, the AFNS model historical estimation precision seems reasonable, although the model is somewhat biased to estimate worse the 0.5-year and 30-year yields as well as the 3-year cap prices than other outputs.

The historical and future state evolution implied by the AFNS model is surprisingly similar to that of the previous case study, see Figures 15 and 8. Although the model parameters in the two cases are close to each other, the estimated historical states could be different due to different model outputs. It seems that an AFNS specified to model the yield curve is a good starting point for cap price modeling, which is encouraging, as cap prices are tightly coupled with yields after all. Given that the state evolution and model parameters are very similar to the previous case study, it is no surprise that the future yield distributions also are. A cap price distribution evolution is demonstrated in Figure 17; the resulting distributions seem convincing, although only a single cap specification is considered. Clearly, more detailed study of future cap prices with different maturities and strikes would be appropriate in more engaged exposure modeling. Additionally, it would be worthwhile to investigate how the cap prices are distributed among different caplets.

The obtained exposure profiles in Figures 19 and 20 seem to be credible given the trade specifications. However, as was already stated in the previous case study, the exposure estimates should be subjected to assessments of validity, error bounds and sensitivity to parameters and model choice. A starting point could be to compare these exposure profiles to ones obtained with an ATSM featuring stochastic volatility.

6 Conclusions

In this thesis, we have shown how affine term structure models (ATSMs) can be applied to quantitative exposure estimation of interest rate derivatives within the counterparty credit risk context. The perspective is application-driven and the purpose has been to show how exposure estimation can be carried out in practice.

The exposure modeling is presented within a framework that follows closely the blueprint of Pykhtin and Zhu (2007). In essence, exposure modeling consists of two distinct computational steps: risk factor simulation and derivative pricing. The purpose of the simulation step is to generate future scenarios for the risk factors. The simulation step should be conducted under the physical probability measure, that is, with a model calibrated to historical observations of relevant quantities. In the pricing step the derivatives considered are valued in each simulated scenario according to the no-arbitrage theory, that is, under the risk neutral probability measure. Distributions characterizing uncertain values of derivatives at future time instances are obtained this way; exposure measures are easy to derive from the future value distributions.

Affine term structure models (ATSMs) is the class of models presented both for the risk factor simulation and derivative pricing. Dynamics of these models can be specified under the physical and risk neutral probability measures, which enables use of ATSMs in both computational steps. The choice to study ATSMs is motivated by their flexible specification and efficient computations that are due to the closed form solutions for the economic quantities of interest.

In order to apply an ATSM, the model parameters need to be estimated. The setup of ATSMs corresponds to that of state space models, which makes it possible to use state space model parameter estimation methods also for ATSMs. At the heart of the parameter estimation problem is the need to evaluate the likelihood of observations given the model and its parameters. For this purpose we have reviewed the continuous-discrete Kalman filter and cubature Kalman filter for linear and nonlinear problems, respectively.

Finally, the different phases of exposure estimation are demonstrated in two case studies. A particular ATSM, the affine arbitrage free Nelson–Siegel model of Christensen *et al.* (2011), is introduced in implementation detail and applied to exposure modeling of an interest rate swap, in the first case study, and a derivative portfolio consisting of an interest rate swap and cap, in the second one. The results are encouraging and suggest that the AFNS model can be applied to exposure modeling of swaps and caps in feasible manner. However, assessing the validity, error bounds and sensitivity of the exposure estimates requires further investigation, which is beyond the scope of this thesis, although essential in any serious exposure modeling endeavor.

The following topics are of interest for future research: exploring exposure estimation with more advanced ATSMs, that feature stochastic volatility, as well as exploring exposure estimation of more complex derivatives, especially swaptions. Additionally, incorporating the multi-curve pricing framework in exposure modeling with ATSMs would be of essential importance. Implications and possible advantages of using prices of derivatives in model calibration is also a topic for further research.

A Appendix

A.1 Solutions 1

Given $\mathbf{\Sigma} = \text{diag}([\sigma_{11}\ \sigma_{22}\ \sigma_{33}])$ and the system

$$
\begin{aligned}
a(t,T,\mathbf{0}) &= \frac{1}{2}\sum_{i=1}^{3}\int_t^T \left(\mathbf{\Sigma}^\top \mathbf{b}(s,T,\mathbf{0})\,\mathbf{b}(s,T,\mathbf{0})^\top \mathbf{\Sigma}\right)_{ii} \mathrm{d}s \\
\mathbf{b}_1(t,T,\mathbf{0}) &= -(T-t) \\
\mathbf{b}_2(t,T,\mathbf{0}) &= -\frac{1}{\lambda}\left(1-e^{-\lambda(T-t)}\right) \\
\mathbf{b}_3(t,T,\mathbf{0}) &= (T-t)e^{-\lambda(T-t)} - \frac{1}{\lambda}\left(1-e^{-\lambda(T-t)}\right),
\end{aligned}
$$

analytical solution for $a(t,T,\mathbf{0})$, that is also provided in Christensen *et al.* (2011), is:

$$
\begin{aligned}
a(t,T,\mathbf{0}) = {} & \sigma_{11}^2\frac{(T-t)^3}{6} \\
& + \sigma_{22}^2\left[\frac{T-t}{2\lambda^2} - \frac{1}{\lambda^3}\left(1-e^{-\lambda(T-t)}\right) + \frac{1}{4\lambda^3}\left(1-e^{-2\lambda(T-t)}\right)\right] \\
& + \sigma_{33}^2\left[\frac{T-t}{2\lambda^2} + \frac{T-t}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{4\lambda}(T-t)^2e^{-2\lambda(T-t)}\right. \\
& \quad -\frac{3(T-t)}{4\lambda^2}e^{-2\lambda(T-t)} - \frac{2}{\lambda^3}\left(1-e^{-\lambda(T-t)}\right) \\
& \quad \left.+\frac{5}{8\lambda^3}\left(1-e^{-2\lambda(T-t)}\right)\right].
\end{aligned}
$$

A.2 Solutions 2

Given $\mathbf{\Sigma} = \text{diag}([\sigma_{11}\ \sigma_{22}\ \sigma_{33}])$ and the system

$$
\begin{aligned}
a(t,T,\mathbf{u}) &= \frac{1}{2}\sum_{i=1}^{3}\int_t^T \left(\mathbf{\Sigma}^\top\,\mathbf{b}(s,T,\mathbf{u})\,\mathbf{b}(s,T,\mathbf{u})^\top\,\mathbf{\Sigma}\right)_{ii}\,\mathrm{d}s \\
\mathbf{b}_1(t,T,u_1) &= -(T-t)+u_1 \\
\mathbf{b}_2(t,T,u_2) &= -\frac{1}{\lambda}\left(1-(1+u_2\lambda)e^{-\lambda(T-t)}\right) \\
\mathbf{b}_3(t,T,u_3) &= \left((T-t)(1+u_2\lambda)+u_3\right)e^{-\lambda(T-t)} - \frac{1}{\lambda}\left(1-e^{-\lambda(T-t)}\right),
\end{aligned}
$$

analytical solution for $a(t,T,\mathbf{u})$ is:

$$
a(t,T,\mathbf{u}) = \frac{1}{\lambda^3}e^{-\lambda(T-t)}a_1 + \frac{1}{24\lambda^3}a_2 + -\frac{1}{8\lambda^3}e^{-2\lambda(T-t)}a_3
$$

where

$$
a_1 = \sigma_{22}^2(1+\lambda u_2) + \sigma_{33}^2(2+\lambda^2(-t+T)u_2 + \lambda(-t+T+u_2+u_3))
$$

$$
\begin{aligned}
a_2 = {} & -4\lambda^3\sigma_{11}^2(t-T)(t^2-2tT+T^2+3tu_1-3Tu_1+3u_1^2) \\
& + 6\sigma_{22}^2[-3+\lambda^2u_2^2-2\lambda(t-T+u_2)] \\
& + 3\sigma_{33}^2\left[-11-2\lambda(2t-2T+2u_2+u_3)+\lambda^2(u_2^2+2u_2u_3+2u_3^2)\right]
\end{aligned}
$$

$$
\begin{aligned}
a_3 = {} & 2(\sigma_{22}+\lambda\sigma_{22}u_2)^2 + \sigma_{33}^2\Big[5+2\lambda^4(t-T)^2u_2^2 \\
& + 2\lambda^3(t-T)u_2(2t-2T-u_2-2u_3) + \lambda(-6t+6T+4u_2+6u_3) \\
& + \lambda^2[2t^2+2T^2+8Tu_2+u_2^2+4Tu_3+2u_2u_3+2u_3^2-4t(T+2u_2+u_3)]\Big]
\end{aligned}
$$

References

Aït-Sahalia, Y. and Kimmel, R. L. (2010). Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions. *Journal of Financial Economics*, 98(1):113–144.

Anderson, B. and Moore, J. (1979). *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, New Jersey.

Arasaratnam, I. (2009). *Cubature Kalman Filtering: Theory & Applications*. PhD thesis, Department of Electrical & Computer Eng., McMaster University.

Arasaratnam, I. and Haykin, S. (2009). Cubature Kalman Filters. *IEEE Transactions on Automatic Control*, 54(6):1254–1269.

Ayyub, B. M. (2003). *Risk Analysis in Engineering and Economics*. CRC Press, Boca Raton, Florida.

Barber, D. (2012). *Bayesian Reasoning and Machine Learning*. Cambridge University Press, Cambridge.

Basel Committee on Banking Supervision (2006). *International Convergence of Capital Measurements and Capital Standards: A Revised Framework – Comprehensive Version*. Bank for International Settlements. http://www.bis.org, Accessed 01-October-2012.

Bianchetti, M. (2010). Two Curves, One Price. *Risk Magazine*, 23(8):66–72.

Black, F. (1976). The Pricing of Commodity Contracts. *Journal of Financial Economics*, 3(1):167–179.

Black, F. and Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *The Journal of Political Economy*, 81:637–654.

Brigo, D. and Mercurio, F. (2006). *Interest Rate Models – Theory and Practice: With Smile, Inflation and Credit*. Springer-Verlag, Berlin Heidelberg.

Canabarro, E. and Duffie, D. (2003). Measuring and Marking Counterparty Risk. In Tilman, L. M., editor, *Asset/Liability Management of Financial Institutions*. Institutional Investor Books, London.

Cesari, G., Aquilina, J., Charpillon, N., Filipovic, Z., Lee, G., and Manda, I. (2009). *Modelling, Pricing, and Hedging Counterparty Credit Exposure*. Springer, Berlin Heidelberg.

Cheridito, P., Filipović, D., and Kimmel, R. L. (2007). Market Price of Risk Specifications for Affine Models: Theory and Evidence. *Journal of Financial Economics*, 83(1):123–170.

Christensen, J. H., Diebold, F. X., and Rudebusch, G. D. (2011). The Affine Arbitrage-Free Class of Nelson–Siegel Term Structure Models. *Journal of Econometrics*, 164(1):4–20.

Christoffersen, P., Dorion, C., Jacobs, K., and Karoui, L. (2012). Non-Linear Kalman Filtering in Affine Term Structure Models. Technical report, CREATES Research Paper. `http://pure.au.dk/portal/files/50944428/rp12_49.pdf`, Accessed 10-Aug-2013.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2):385–407.

Dai, Q. and Singleton, K. J. (2000). Specification Analysis of Affine Term Structure Models. *The Journal of Finance*, 55(5):1943–1978.

De Prisco, B. and Rosen, D. (2005). Modelling Stochastic Counterparty Credit Exposures for Derivatives Portfolios. In Pykhtin, M., editor, *Counterparty Credit Risk Modelling*. Risk Books, London.

Dempster, M. A., Medova, E. A., and Villaverde, M. (2010). Long-Term Interest Rates and Consol Bond Valuation. *Journal of Asset Management*, 11(2):113–135.

Diebold, F. X. and Li, C. (2006). Forecasting the Term Structure of Government Bond Yields. *Journal of Econometrics*, 130(2):337–364.

Duarte, J. (2004). Evaluating an Alternative Risk Preference in Affine Term Structure Models. *Review of Financial Studies*, 17(2):379–404.

Duffee, G. R. (2002). Term Premia and Interest Rate Forecasts in Affine Models. *The Journal of Finance*, 57(1):405–443.

Duffie, D., Filipovic, D., and Schachermayer, W. (2002). Affine Processes and Application in Finance. Working paper, National Bureau of Economic Research Cambridge, Mass., USA. `http://www.nber.org/papers/t0281.pdf`, Accessed 13-May-2013.

Duffie, D. and Kan, R. (1996). A Yield-Factor Model of Interest Rates. *Mathematical Finance*, 6(4):379–406.

Duffie, D., Pan, J., and Singleton, K. (2000). Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica*, 68(6):1343–1376.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford University Press, New York.

Eisenführ, F., Weber, M., and Langer, T. (2010). *Rational Decision Making*. Springer, Berlin Heidelberg.

Feldhütter, P. (2008). Can Affine Models Match the Moments in Bond Yields? Working paper, Copenhagen Business School. `http://www.feldhutter.com/RiskPremiumPaper.pdf`, Accessed 29-May-2013.

Feller, W. (1951). Two Singular Diffusion Problems. *Annals of Mathematics*, 54:173–182.

French, S. (1986). *Decision Theory: An Introduction to the Mathematics of Rationality*. Halsted Press, New York.

Gibson, M. S. (2005). Measuring Counterparty Credit Exposure to a Margined Counterparty. In Pykhtin, M., editor, *Counterparty Credit Risk Modelling*. Risk Books, London.

Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering*. Applications of Mathematics. Springer, New York.

Gregory, J. (2010). *Counterparty Credit Risk: The New Challenge for Global Financial Markets*. John Wiley & Sons, Chichester, United Kingdom.

Grewal, M. S. and Andrews, A. P. (2001). *Kalman Filtering: Theory and Practice Using MATLAB*. Wiley-Interscience, New York.

Harrison, J. M. and Kreps, D. M. (1979). Martingales and Arbitrage in Multiperiod Securities Markets. *Journal of Economic Theory*, 20(3):381–408.

Harrison, J. M. and Pliska, S. R. (1981). Martingales and Stochastic Integrals in the Theory of Continuous Trading. *Stochastic Processes and Their Applications*, 11(3):215–260.

Hull, J. C. (2011). *Options, Futures, and Other Derivatives*. Prentice Hall, Upper Saddle River, New Jersey, 8th edition.

Ito, K. and Xiong, K. (2000). Gaussian Filters for Nonlinear Filtering Problems. *IEEE Transactions on Automatic Control*, 45(5):910–927.

Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*. Academic Press, New York.

Jong, F. d. (2000). Time Series and Cross-Section Information in Affine Term-Structure Models. *Journal of Business & Economic Statistics*, 18(3):300–314.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45.

Karatzas, I. A. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York.

Klebaner, F. C. (2005). *Introduction to Stochastic Calculus With Applications*. Imperial College Press, London.

Kloeden, P. E. and Platen, E. (1999). *Numerical Solution to Stochastic Differential Equations*. Springer, Berlin Heidelberg New York.

Lund, J. (1997). Non-Linear Kalman Filtering Techniques for Term-Structure Models. Working paper, Aarhus School of Business. `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.55.8535&rep=rep1&type=pdf`, Accessed 10-Aug-2013.

Maruyama, G. (1955). Continuous Markov Processes and Stochastic Equations. *Rendiconti del Circolo Matematico di Palermo*, 4(1):48–90.

Mbalawata, I. S., Särkkä, S., and Haario, H. (2013). Parameter Estimation in Stochastic Differential Equations with Markov Chain Monte Carlo and Non-Linear Kalman Filtering. *Computational Statistics*, 28(3):1195–1223.

McNeil, A., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton Series in Finance. Princeton University Press, Princeton, New Jersey.

Modarres, M. (2006). *Risk Analysis in Engineering*. Taylor & Francis, Boca Raton, Florida.

Nelson, C. R. and Siegel, A. F. (1987). Parsimonious Modeling of Yield Curves. *Journal of Business*, 6:473–489.

Øksendal, B. (2003). *Stochastic Differential Equations: An Introduction with Applications*. Springer, Heidelberg New York, 6th edition.

Pallavicini, A. and Tarenghi, M. (2010). Interest-Rate Modeling With Multiple Yield Curves. Working paper, SSRN. `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1629688`, Accessed 01-May-2013.

Pykhtin, M. (2005). *Counterparty Credit Risk Modelling*. Risk Books, London.

Pykhtin, M. (2009). Modeling Credit Exposure for Collateralized Counterparties. *Journal of Credit Risk*, 5:3–27.

Pykhtin, M. and Zhu, S. H. (2007). A Guide to Modeling Counterparty Credit Risk. *GARP Risk Review*, (July/August):16–22.

Rebonato, R. (1996). *Interest-Rate Option Models*. John Wiley & Sons, New York.

Rebonato, R., Mahal, S., Joshi, M., Bucholz, L.-D., and Nyholm, K. (2005). Evolving Yield Curves in the Real-World Measures: A Semi-Parametric Approach. *Journal of Risk*, 7(3):29–62.

Särkkä, S. (2006). *Recursive Bayesian Inference on Stochastic Differential Equations*. Doctoral dissertation, Helsinki University of Technology.

Särkkä, S. and Solin, A. (2012). On Continuous-Discrete Cubature Kalman Filtering. In *Proceedings of SYSID*, page 1210–1215.

Schrager, D. F. and Pelsser, A. A. (2006). Pricing Swaptions and Coupon Bond Options in Affine Term Structure Models. *Mathematical Finance*, 16(4):673–694.

Shreve, S. (2004). *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer, New York.

Singleton, K. J. (2006). *Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment*. Princeton University Press, Princeton, New Jersey.

Singleton, K. J. and Umantsev, L. (2002). Pricing Coupon-Bond Options and swaptions in Affine Term Structure Models. *Mathematical Finance*, 12(4):427–446.

Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the Theory of the Brownian Motion. *Physical Review Online Archive (Prola)*, 36(5):823–841.

Vasicek, O. (1977). An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*, 5(2):177–188.

Williams, D. (1991). *Probability with Martingales*. Cambridge University Press, Cambridge.

Wu, Y., Hu, D., Wu, M., and Hu, X. (2006). A Numerical-Integration Perspective on Gaussian Filters. *IEEE Transactions on Signal Processing*, 54(8):2910–2921.

Zhu, S. and Pykhtin, M. (2006). Measuring Counterparty Credit Risk for Trading Products Under Basel II. In Ong, M. K., editor, *The Basel Handbook: A Guide for Financial Practitioners*. Risk Books, London, 2nd edition.